University of Pennsylvania

TRANSFORMATIONS AND DISCOURSE ANALYSIS PAPERS

28. Problems in the Computation of English Sentence Structure (Substring Analysis)

1963

Naomi Sager

Contents

1. String identifier

For purposes of computation, one of the most useful statements of a string grammar is that a sentence formula is decomposable into strings without residue. This means that if one member of a string can be identified in the sentence formula, then the other members must also be there. Suppose a particular string contains one symbol which is unique to that string, e.g. K_6 (*whoever*) is unique to the string $K_6\ V^+$ $(\underset{K_6}{\underline{\text{whoever}}}\ \underset{V}{\underline{\text{reads}}}\ \underset{+}{\underline{\text{Rimbaud}}})$; if we can locate K_6 in a sentence formula, we can immediately impose a grammatical requirement, namely: V^+ must also be locateable in the sentence formula. More broadly speaking, if each string has an identifying symbol we can separate the problem of identifying a string's occurrence in a sentence formula (discovering the fact of its occurrence), from that of computing its structure. It might be possible for example to identify which strings occur in a sentence formula by a quick scan for "string identifiers" (assuming each string has one) before analyzing the structure of the sentence. One might then choose to analyze particular portions of the sentence formula first, e.g. parts relatively free of ambiguity, or to use different methods for different portions.

Many strings in English have what might be called a string identifier (defined below), e.g. K_4 (*what*), while not unique to a single string, is a convenient identifier of $K_4\ V^+$ ($\underset{K_4}{\underline{\text{what}}}\ \underset{V}{\underline{\text{succeeds}}}$) and of $K_4\ N\ V^-$ ($\underset{K_4}{\underline{\text{what}}}\ \underset{N}{\underline{\text{they}}}\ \underset{V^-}{\underline{\text{said}}}$) because K_4 occurs only as a member of one of these two strings or in a question, whereas the other members of these two strings V^+, N, V^-, occur in many strings. Let us say that the symbol X is an identifier of the string if X is a member of α and X has a much more limited distribution than other members of α.

Do all the strings of English have identifiers, and if so, where is the identifier located in the string? A classification of the strings of English listed in TDAP 27 (pp. 37-41), about 100 in number, yields

a) the main body of strings - which do have identifiers;

b) object strings (about 30) - which have a modified type of identifier;

c) one symbol strings (about 10) - which serve as their own identifiers;

d) certain strings which begin with N (13 in number) - which do not have identifiers.

a) All of the strings in what we have called the main body of strings have identifiers at a stateable distance from the leftmost element. We find that most strings have identifiers which are the leftmost symbol of the string, for example the string $\text{ç}_2\ N\ V$ (<u>when it reaches us</u>). Two groups of strings are exceptions; the "P K" strings in which the identifier is the second symbol from the left, and the "N P K" strings in which the identifier is the third symbol from the left.[1] We have:

---

[1]But see section 2 below on recursion in P K and N P K strings.

$\downarrow$

$P\ K_2\ N\ V^{P^2}$	e.g. to whom it may be problematic
$P\ K_3\ N\ V^P$	e.g. beyond which I can't see
$P\ K_3\ N\ N\ V^P$	e.g. at which position one is rejected
$P\ K_4\ N\ V^P$	e.g. of what it consists
$P\ K_5\ N\ N\ V^P$	e.g. in whose laboratory they worked
$P\ K_7\ N\ V^P$	e.g. on whomever one relies
$P\ K_8\ N\ N\ V^P$	e.g. on whatever page it is
$P\ K_9\ N\ N\ V^P$	e.g. to whatever length it extends
$P\ K_{10}\ N\ N\ V^P$	e.g. on whosever word you take it

$\downarrow$

$N\ P\ K_3\ V+$	e.g. some of which came to light
$N\ P\ K_3\ N\ V^-$	e.g. a sample of which we obtained
$N\ P\ K_5\ N\ V^+$	e.g. a part of whose work repeats ours
$N\ P\ K_5\ N\ N\ V^-$	e.g. a portion of whose work we saw
$N\ P\ K_2\ V^+$	e.g. some of whom stayed to dinner
$N\ P\ K_2\ N\ V^-$	e.g. some of whom we invited for dinner

---

[2]Superscripts P, +, -, e.g. V^P, concern the computation of the object string, and are defined in section 2 below.

b) The object strings do not contain an identifier as defined above. However, the selection of object strings which can occur following a particular verb V_i in the sentence formula is a property of that verb, i.e. V_i, which is not a member of the object strings which it specifies, functions like a string identifier in that it may be used to set up the requirements that certain other elements must also occur in the sentence formula. Like K_4, which is an identifier of more than one string, V_i may specify more than one possible object string, only one of which actually occurs; the requirement is that once V_i occurs, one of the possible object strings of V_i must occur.

c) One symbol strings are their own identifiers since in general they are not symbols which occur in other strings (except N of compound noun, N of apposition, discussed below). The 1 symbol strings are:

Symbol (class name)	Description of class	Type of string	Word example
A	adjective	left adjunct of N or N's	free man
S	past participle	" " " " " "	freed man
G	verb with ing	" " " " " "	singing sparrow
Z	color word	" " " ", replacement of N	red flowers, red is my favorite
Q	quantifier including numbers	" " " ", replacement of N	some flowers, many protested
N	of cpd nound	" " "	brick wall
N's	possessive	" " "	boy's mother
T	article class	non-repeatable left element of N or N's	a chair
B	this, these (=T A)		this chair
R	pronoun	replacement of N	we
N	apposition	right adjunct of N	his father, a lawyer
&	unary conj.	sentence adjunct	however
W	certain verb aux.	left adjunct of V	will, may, can
D	adverb	left adjunct of A, S, G, Z, Q	partly free, lately freed, definitely coming, very red very few
		" " " D	very thoroughly, completely unashamedly
		" " " P	completely under the covers
		" " " V	we definitely want to go
		right " " V	we want strongly that he accept
		sentence adjunct	generally, the items are found in groups

In addition there are subclassifications of some of the above which occur in definite positions, e.g.

Q_t a subclass of quantifiers which can occur to the left of T (many the man)

D_t a subclass of adverbs which can occur to the left of T or Q_t (just a man)

A_v a subclass of adjectives which can occur to the right of of N (the people present)

and various subclasses of T.

Particular words may have more than one classification, e.g. more D/Q (We liked him more, more girls) one Q/R (one chair broke, one would like to know), many Q_t (The many persons, many the man) There is need for more work on the classification of articles and quantifiers and adverbs, including no and not

d) A number of strings which begin with N do not have string identifiers as defined.

Symbol	Type of String	Word Example
N	left portion of cpd N	the baseball scores
N {A, S, G}		a book buying spree
N G	replacement strings of N	elephant hunting is ...
$N\ G^{+}$		people buying souvenirs is ...
$N\ S^{+}$	sentence adjuncts based on N be^{+}	his mission accomplished, he ...
N A		the room free, we ...
N D		the landlord out, the tenant ...
N PN		the ship on its way, we ...
N N		the defendant a known SS man, the trial
$N\ G^{+}$		the book selling, the publishers
N (apposition)	right adjuncts of N	his son, a painter
N V		the book he bought
$N\ V^{+}$[1]		the reason we are here
$N\ V^{-}_{14}$[2]	sentence adjuncts	he came, we think
N_7		today

[1] Only certain nouns, like reason, fact, can be adjoined by an $N\ V^{+}$ string.

[2] V_{14} is a class of verbs, such as think, suppose, which may have as object a complete sentence.

Since the symbol N occurs in many strings it is not by itself useful in announcing the presence of a particular N-containing string in the sentence formula. The same is true for most of the other symbols in the strings listed above. Thus these strings have no identifier as we have defined it. However, these strings are divided as to syntactic function, i.e. as to the positions they can occupy, and there are other restrictions which help to distinguish them from other sequences that include N. These strings discourage a computational scheme which first scans for string boundaries (using identifiers as cues) and then computes structure; they favor a scheme which brings a string in relation to other strings at the same time it identifies the string.

Summary: The main body of strings can be identified by a symbol which is either the leftmost element of the string or is in a stateable position relative to the leftmost symbol. The presence of an object string in the sentence formula is indicated by a verb which occurs to the left of the object string, in particular the verb of the string which includes the object string. Thus the great bulk of strings are left-oriented. The one symbol strings are neutral. There is also a group of about 15 strings which begin with N, the identification of which presents a problem in that they contain no symbol which marks them off from other strings.

2. Omission

2.1 Verb Object and Omission:

In analyzing a sentence formula it is necessary to associate an occurring object string with its verb. In TDAP 27 this was accomplished as follows:

Preliminary: We assume that the possible object strings of each English verb V_i are listed in a dictionary prior to computation.

1. Object specification: Upon encountering V_i in a sentence formula, copy its possible object strings from the dictionary into a register (called object register).

2. Comparison: Compare, symbol by symbol, each object string in the register with the sequence of symbols following V_i in the sentence formula. If the symbols of an object string O_i are identical with the symbols following V_i in the sentence formula, the association of O_i with V_i is a possibly correct analysis.[1]

   NOTE: We say "possibly correct analysis" because O_j (a shorter object string having the same symbols as the first portion of O_i) may also be a correct analysis and the decision as to whether the object of V_i is O_i or O_j or both (permanent ambiguity) depends on the overall analysis of the sentence formula.

Depending on the string which contains a verb, the verb's object strings occur in slightly different form. In particular, if one adopts as a reference form of an object string, its form when its associated verb is the main verb

[1]In an actual computational scheme, such as the one discussed in TDAP 27, there is provision for the occurrence of adjunct strings among the symbols of the object string.

of a sentence (i.e. of a center string), one finds that when the same verb occurs in certain non-center strings, the object string occurs with one less N. For example, the verb attribute has for object string (reference form) $N\ P_{to}\ N$ e.g. We attributed $\underset{N_1}{\text{our success}}$ $\underset{P_{to}}{\text{to}}$ $\underset{N_2}{\text{chance}}$; employing a which clause we might have: Our success which we attributed $\underset{P_{to}}{\text{to}}$ $\underset{N_2}{\text{chance}}$...or $\underset{N_2}{\text{Chance}}$, which we attributed $\underset{N_1}{\text{our success}}$ $\underset{P_{to}}{\text{to}}$... In each of the latter cases, where the verb attribute occurs in a which string, we find following the verb attribute the object string $N\ P_{to}\ N$ minus one N.[1]

One may ask: Suppose a verb object string contains more than one N, such as the N P N object string of the verb attribute; can each N independently absent? The answer for most object strings is yes as the above example with attribute illustrates. We also find that when an object string contains a verb, the omission of N can be "passed along" to the object strings of the contained verb. E.g. in

The guide we told the children to ask the visitors to wait for was late.

we have an $N\ V^-$ string (beginning with we told), but it is not the object string of told which occurs with one less N than in its reference form; rather we find a chain of three object strings after told, the N being omitted from the last of the three.

---

[1] We indicate the fact that an N is absent ("omitted") from the reference form of the object string of a V when that V occurs in a particular string by writing V^- for the V in the string. By contrast, V^+ stands for a verb having its object string in reference form. Thus the which string used in the example above is written $K_3\ N\ V^-$ and the declarative center string is written $N\ V^+$.

Verb		Object String (Ref. form)	Omission
V_1	(told)	N to V_2^+ (children to ask)	"passed along"
V_2	(ask)	N to V_3^+ (visitors to wait)	"passed along"
V_3	(wait)	P N (for the guide)	P N $\rightarrow$ P

We can now formulate a rule for obtaining from the reference form of the object strings of V_i, the form of these object strings when V_i occurs as the V^- of a string. Let the reference form of the object strings of V_i be in an object register. Let the "-" of V^- indicate that the Rule of Omission is to be applied.

Rule of Omission: 1. For each N_j in the object register, rewrite the string which contains N_j, leaving out N_j.

2. For each object string O_k (in reference form) which contains a verb V_k^+, add to the strings in the register a copy of O_k with V_k^+ rewritten as V_k^-.

Part I of the Rule governs the omission itself; part 2 provides for "passing along" the omission.

Example of the Application of the Rule of Omission

We take a sample verb V_i (told) in Column 1. We show the object strings of V_i in reference form, i.e. the contents of the Object Register before the Rule of Omission (R. of O.) is applied, in column 2. We show the results of applying R. of O., i.e. the contents of the Object Register after R. of O. is applied in column 3. Columns 4 and 5 are devoted to word illustrations of the forms in column 3 as follows: On each line, we let the verb *told* correspond to the V^- of an N V^- string, *we told*.... If the N-omission is according to part 1 of R. of O., a noun from the object string of *told* is omitted (column 4), e.g. in row 1, column 4 *children* is omitted from *we told (the children) to invite the visitors*. The omitted noun appears in front of the N V^- string. If the N-omission is according to part 2 of R. of O. it means the signal for omission has been "passed along" to a verb, e.g. *invite*, contained in an object string of *told* (column 5), e.g. in row 1, column 5, *visitors* is omitted from *we told the children to invite the visitors*. The place of the omitted N is indicated by ( ).

1 Verb	2 Object Strings (contents of object register before rule is applied)	3 Rule of Omission applied (contents of Object Register after Rule is applied)	4 Word Example Part 1 of Rule of Omission	5 Word Example Part 2 of Rule of Omission
told	N to V^+	to V^+	the children we told ( ) to invite the visitors	
		N to V^-		The visitors we told the children to invite ( )...
	N P_{of}N	N P_{of}	the adventures we told him of ( )...	
		P_{of}N	the man we told ( ) of our adventures	
	N P_{about}N	N P_{about}	the adventures we told him about ( )	
		P_{about}N	the man we told ( ) about our adventures	
	N N_8	N	the tale we told ( ) the man	
		N_8	the man we told ( ) the tale	
	N	φ	the man we told ( )	
	N_8	φ	the tale we told ( )	

At the beginning of this section we described a procedure for computing object strings in their reference form, i.e. objects of V^+ . Now suppose V_i occurs in the V^- position of a string, e.g. as the verb in the string $N\ V^-$, we can adpat the procedure for computing the object of V^+ for use with objects of V^- by adding to step 1 (specification) the following:

Apply the Rule of Omission to the object strings in the object register.

Table 1 below shows the application of the Rule of Omission to the objects strings of English. Exceptions to the rule are noted by footnotes at the end. Columns 1, 2, 3 of the table are similar to columns 2, 4, 5 respectively, of the Example above. In addition, in column 1 of Table 1, the N-to-be-omitted is underlined. In some cases, another string in place of $N\ V^-$ is used in constructing the word illustrations of columns 2 and 3.

TABLE 1

Omission in Verb Object Strings

1 Object String (ref. form)	2 Example of N-omission	3 Example of N-omission passed along 1 step
$\underline{N}$	people whom we found ( )	
$\underline{N_{pl}}$	the facts we correlate (N_{pl})	
$\underline{N\ \text{and}\ N}$[1]	this and that we correlate (N & N)	
$\underline{N}$ with $\underline{N}$	the x we correlate (N_1) with y the y we correlate x with (N_2)	
$\underline{V_n}$	the dream he dreamt ( )	
N $\underline{N}$	the book we gave him ( )	
$\underline{N}\ P_{to}\ \underline{N}$	the book we gave (N_1) to him	
P $\underline{N}$	the man we rely on ( )	
$\underline{N}\ D_P$	the game we broke ( ) up	

TABLE 1 (continued)

	1 Object String (ref. form)	2 Example of N-omission	3 Example of N-omission passed along 1 step
	$D_P \ \underline{N}$	the game we broke up ( )	
	G^{+}		stories we like reading ( )
	$P_i \ G^{+}$		the passage he refrained from quoting ( )
to	V_o^{+}		the children we want to teach ( )
(it is)	$A_8 \ H \ \underline{N} \ V^{+2,4}$	the consequences it is possible ( ) will result	the course it is possible that none will take ( )
(it is)	A_8 $\underline{\text{for N}}$ to $V^{+3,4}$	the person for whom it is possible (for N) to speak like that	the cake it is possible for me to bake ( )
(it is)	A_8 to V_o $+4$		the music it is possible to compose ( )
	$A \ P \ G^{+}$		The music he is good at composing ( )
	$\underline{N} \ V^{+}$	the machine we made ( ) work	the load we made the horse carry ( )
	$(H) \ \underline{N} \ V^{+12}$	the experiment we thought ( ) worked	the flowers we hoped that he would plant ( )
	$(H) \ N \ V_{\phi}^{21}$	the book we insisted ( ) appear	the flowers we preferred he plant ( )
	$\underline{N} \ A$	the man we thought ( ) wise	
	$\underline{N} \ D_2$	the man we thought ( ) here	
	$\underline{N} \ N^{5}$	the man we thought ( ) an Englishman	
	$\underline{N} \ P \ \underline{N}$	the man we thought (N_1) in Philadelphia the city we thought the man in (N_1)	

TABLE 1 (continued)

1 Object String (ref. form)	2 Example of N-omission	3 Example of N-omission passed along 1 step
N as D_2	the man we consider ( ) as here	
N as G^+	the man we consider ( ) as being fair	the battles we consider him as having won ( )
N as N^5	the man we consider ( ) as candidate	
N G^+	the classes we opposed ( ) starting at 8 A.M.	the classes which we oppose starting ( ) at 8
N H N V^+	the man we told (N_1) a risk would result	
	the riot we told the man (N_2) would result	the man we told them nothing could stop ( )
N P_i G^+	the man we restrained ( ) from going	certain rules we restrained them from breaking ( )
N to V_o^+	the plans we want ( ) to succeed	the conditions we want the plans to alleviate ( )
N H N should V^+	the man we told ( ) that he should go home	
for N to V_o^{+6}		the load we prefer for him to carry ( )
P N H N V_o^{+7}		
A P N^8		

[1] N and N are to be treated as a single N

[2] H is dropped whenever the adjacent N is omitted from the object string.

[3] For = must be moved along with the N

[4] This object string is a possible object of the verb be only when the subject of be is it.

[5] the second N is not omittable.

[6] The N following for is not omittable.

[7] There are no omissions from the object string P N H N V_o^+ e.g. ask of him that he grease the car.

[8] The N is not omittable from A P N of the type quick to anger.

2.2 A more general picture of "N-omission"

Consider:

The book exists	The book which exists
He found the book	The book which he found
He selected the title for the book	The book which he selected the title for

It is apparent that a similar relation exists between the left and right utterance of each pair. The sentence on the left appears on the right following the word <u>which</u> with <u>the book</u> omitted. If we compare the grammatical role of <u>the book</u> in each of the three sentences on the left we find that it is respectively: the subject, the object of the verb, the noun of a prepositional phrase adjoined to the object of the verb.

Up until now we have considered N-omission only of the type found in the second example, namely, omission of N from the object. The reason is that we can state the structure of strings with other kinds of N-omission in advance of computation without reference to what occurs in a particular sentence formula, whereas in omission of N from the verb object we must wait until a particular verb is encountered in the sentence formula in order to obtain the structures from which to omit an N. Thus, if we want to maximize the stored information in place of deriving expected structures as we go along, we separate the object case from all the others, and define "N-omission" operationally only in relation to strings which contain V^{-} , as was done in 2.1 above.

But for some purposes, it is worthwhile to consider N-omission in a broader sense, covering all three cases above, and others, regardless of how the information is arranged for a string computation. For one thing, it is helpful in understanding how the many strings listed in a string grammar derive from a few basic forms; in particular, in the analysis used here, from a single $N\ V^+$ center string. Further, when arranged in tabular form, it is easier to check how much of English is covered and in what way by the string grammar.

Therefore, we add below, a table of strings which involve N-omission (in the broader sense) indicating the grammatical role of the omitted N with reference to an $N\ V^+$ string and its adjuncts. As an example of how to read the table consider the row which begins with K_3 which. The check in the intersection of the K_3 row with the V^+ column indicates that the string $K_3\ V^+$ exists in English. A word example of $K_3\ V^+$ is obtained by reading the word which, followed by the word example V^+, exists, i.e. which exists. Similarly, the check in the intersection of K_3 row with the $N\ V^-$ column indicates that the string $K_3\ N\ V^-$ exists in English and the example is which he found. Each column specifies the source of the N-omission, so that the original $N\ V^+$ with its adjuncts can be reconstructed. E.g. the column V^+ was obtained by omitting the subject N from $N\ V^+$; if we supply a noun (the book) in subject position, we obtain the book exists.

TABLE 2a

Omission of N in Strings

1	2	3	4	5	6	7
Omission of:	Subject N	N of Object	N of PN; PN adjunct of V or of object N	Adjective of subject N	Adjective of object N	Adjective of N of PN; PN adjunct of verb of of object N
String (minus string identifier):	V^+ e.g. exists	N V$-$ e.g. he found	N V^+ P e.g. he selected the title for	N V^+ e.g. estate survives	N N V e.g. house they choose	N N V^+ P e.g. house we held the meeting in
String identifier						
φ		x	x			
H (that)	x	x	x			
K_1 (who)	x					
K_2 (whom)		x	x			
K_3 (which)	x	x	x	x	x	x
K_4 (what)	x	x	x	x	x	x
K_5 (whose)				x	x	x
K_6 (whoever)	x					
K_7 (whomever		x	x			
K_8 (whichever	x	x	x	x	x	x
K_9 (whatever)	x	x	x	x	x	x
K_{10} (whosever				x	x	x

TABLE 2b

Omission of:	PN adjoined to subject N	PN adjoined to object N	Adjective of N of PN adjoined to subject N	Adjective of N of PN adjoined to N of object
String (minus initial portion through K)	V^+ e.g. exists	$N\ V^-$ e.g. he found	$N\ V^+$ e.g. estate survives	$N\ N\ V^-$ e.g. house they chose
Initial portion of string				
$N\ P\ K_2$ e.g. an heir of whom	x	x		
$N\ P\ K_3$ e.g. an heir of which	x	x		
$N\ P\ K_5$ e.g. an heir of whose			x	x

Omission may be "passed along" to adjuncts included in a K-string, particularly PN adjuncts. Thus column 4 and 7 of Table 2a and all of Table 2b show only the first of a set of recursively extendable forms with omission.

For example, in Table 2a:

column 4 extended

$K_3\ N\ V^+\ P$	which he selected the jacket for
$K_3\ N\ V^+\ P\ N\ P$	which he selected the picture on the jacket for
$K_3\ N\ V^+\ P\ N\ P\ N\ P$	which he selected the colors in the picture on the jacket for

column 7 extended (more awkward)

$K_5\ N\ N\ V^+\ P$	whose house we held the meeting in
$K_5\ N\ N\ V^+\ P\ N\ P$	whose " " " " " " the foyer of

Examples from table 2b are:

Column 3 extended

$N\ P\ K_3\ N\ V^-$	(the estate), an heir of which he found
$N\ PN\ P\ K_3\ N\ V^-$	" " , a grandson of an heir of which he found

column 4 extended

$N\ P\ K_5\ N\ V^+$	(the man), an heir of whose estate survives
$N\ P\ N\ P\ K_5\ N\ V^+$	" " , a grandson of an heir of whose estate survives

In all cases the omission is from the last P N of a string of P Ns.

The K strings in columns 4 and 7 of Table 2a (not ϕ, H) may also appear with P preceding K instead of P at the end of the string, e. g. $P\ K_3\ N\ V^+$ (for which he selected a title). This is one source of P K strings. The other source of P K strings is columns 3 and 6 when the object string from which N is to be omitted contains P N. In that case, P N is omitted from the reference form of the object string and the P "is moved" in front of K, e.g. $P\ K_3\ N\ V^-$(on whom he relied).[1]

In the case of omission of an adjective (more properly, adjectival element, see Section 4 below), the adjectival element is often N's (John's). Should the N's in question have its own adjectives, these are carried along with the N's; they are omitted from the string and appear in front of the K with the N.

---

[1]In TDAP 27 an instruction OBJ P gives the procedure for computing the object of the verb (V^P) of a PK string. It is similar to the procedure described in 2.1 above for computing the object of V^- . But before adjusting the object strings in the object register OBJ P institutes a comparison of the head P (PK) with the Ps in the object register to determine whether the object string is of the form of columns 4 and 7 or of the form of columns 3 and 6.

E. g. N_1's N_2's N_3 (The $\underset{1}{\text{man's}}$ $\underset{2}{\text{hats'}}$ $\underset{3}{\text{feather}}$) →

a) N_1 K_5 N_2's N_3 . . . . (The $\underset{1}{\text{man}}$ whose $\underset{2}{\text{hat's}}$ $\underset{3}{\text{feather}}$ . . .)

[Here there are no adjectives of N_1 to be carried.]

b) N_1's N_2 K_5 N_3. . . (The $\underset{1}{\text{man's}}$ $\underset{2}{\text{hat}}$ whose $\underset{3}{\text{feather}}$ . . .)

[The adjective N_2 is carried along].

3. Conjunctions

When the element C (and, or, but[1]) occurs in the sentence formula, we can find a structure to the left of C (α) and one to the right of C (α') which bear to each other a certain relation which we will call C-equivalence (α C α). This relation is first stated for particular cases and then a general form for C-equivalences in strings is given.

3.1 C-Equivalence of left adjuncts of N: As a preliminary, we name the main left adjuncts of N. These are:

---

[1]The element than (preceded in the sentence formula by more or less or rather or worse or -er) and the element as (preceded in the sentence formula by as) can be classed in C. Discussion of these is temporarily deferred.

Pre-article quantifier	Q_t		e.g. many
Article	T		
		T_1 a	
		T_2 the	
		T_3 some	
		T_4 either, neither	
	B	(this, that, these, those)	
Quantifier	Q		e.g. more, many, few, several
Adjectival element			
	$\bar{A}$	A	e.g. young
		G	e.g. gathering
		S	e.g. given, determined
		Z	e.g. blue
		N's	e.g. John's, his
Left portion of compound noun			
		N	e.g. telephone in telephone repairman
		N $\bar{A}$	e.g. friction generating to friction generating elements

When C occurs in a sentence formula immediately following a left adjunct of N -- call it X --, the most common C-equivalence is X C X', i.e. two elements of the same class conjoined by C. X C X' will be followed by an N of which both X's are adjuncts, with X and X' either grouped together into one adjunct or adjoined independently (see below); other adjuncts may intervene between X C X' and N. Explicitly, thwn C occurs among left adjuncts of N, we have the C-equivalences:[1]

---

[1] In the examples, the words corresponding to the C-equivalence are underlined.

(1)	B C B'	this and / or / but (not) that free element
(2)	Q C Q'	more and more people
(3)	$\bar{A}$ C $\bar{A}$'	young and talented candidates
(4)	Z C Z'	blue or green walls
(5)	N's C N's'	John's or my considered opinion
(6)	N C N'	telephone and telegraph repairmen
(7)	N$\bar{A}$ C N$\bar{A}$'	tin plated and air free vessel

Let X, Y represent two different left adjuncts of N; C-equivalences of the form X C Y' occur more rarely than X C X'. A few examples of varying likelihood of occurrence are:

Z C $\bar{A}$'	blue and fresh sky
Q C $\bar{A}$'	more and good cooks
N's C Z'	John's and blue wall
N's C $\bar{A}$'	John's but false opinion
N C N$\bar{A}$'	glass and air free vessel
N$\bar{A}$ C N'	mountain climbing and pack mules.

X C Y' is more likely when X and Y are 'neighbors' in the ordered list of left adjuncts given at the beginning of the section, e.g.

$\bar{A}$ C Z'	cloudy and grey sky
Z C $\bar{A}$'	grey and cloudy sky

are more likely than

Q C Z'	several and grey clouds
Z C Q'	grey and several clouds

Also X C Y' is more likely when Y is preceded by an adjunct, e.g.

Z C $\bar{A}'$ — a blue and remarkably fresh sky

Q C $\bar{A}'$ — more and, in my opinion, good cooks are needed.

(Compare with blue and fresh sky, more and good cooks).

The last two examples also show that C-equivalent structures α and α' are not always contiguous with C; for instance, the structure immediately following C may be a left adjunct of α' or a sentence adjunct. The sole left adjunct of left adjuncts of N is D (adverb), so that when the element after C is D, either α C α' = D C D' or else the D which follows C is a left adjunct of the α' structure, e.g.

(8) D C D' — discernably or allegedly studious members of the class

A C A' — lively and allegedly studious members of the class

An ambiguity arises when it is α , the left member of the C-equivalence, which has left adjunct D. Then we do not know whether D is an adjunct of α alone or of both α and α' , e.g. very in

very lively and studious members of the class.
D A C A

In general, if α and α' can both have a particular left adjunct, and that adjunct occurs to the left of α in the sentence formula, we do not know whether to associate the adjunct with α or with α and α' . This type of structurally undecidable ambiguity is also found where right adjuncts of α or α' occur to the right of α' , e.g. yesterday in

The experiment which we started and they finished yesterday.

Another ambiguity concerns whether α C α' should be grouped as a unit or not, e.g.

Telephone and telegraph repairmen

( ) = men who repair both telephones and telegraph equipment.

( ) ( ) = a group, some of whom are telephone repairmen and some telegraph repairmen.

Sometimes the subclassification of the Ns is such as to rule out a grouping, e.g. there is only one grouping for: electrician and telephone repairman.

When a sequence of left adjuncts of N begins with T, it is possible to obtain C-equivalent structures α , α' which are sequences of several elements, each sequence beginning with (the same) T, (But not T C T) e.g.

TA C TA   a large and a publicly owned garden

although more frequently the second occurrence of the article is absent. Equivalences such as TA C TA where α includes more than one symbol as well as all the other equivalence of left adjuncts of N_1 are obtained automatically from the general form of Type 1 C-equivalence of strings (3.4, 3.5 below) by considering the nounphrase headed by N as a string.

3.2 C-equivalences of Ns and N-replacement Strings

When C occurs in the sentence formula after N (where N is not a left portion of a compound N) we have the C-equivalence:

(1) N C N'   dawn and twilight

There may be left adjuncts of N' occurring between C and N', e.g.

chocolate and hard candies. There are several questions as to the grouping of N and N' and their adjuncts. Sometimes N' can stand as a separately grouped element (with or without adjuncts), e.g.

A (N) C (N') mechanical toys and television;

sometimes it must be grouped with N,

(a) because N' is a count-noun and requires the article which occurs to the left of N, e.g.

T (N C N) a table and chair[2]

(b) because the two Ns have a particular relation, e.g.

A (N C N) the mixed butter and sugar

According to Vendler, there is an ordered subclassification of adjectives (extendable to all left adjuncts of N),[1] such that if a member of each subclass occurs in a sequence, the order of the sequence is fixed. If any subclass is not represented in the sequence its "slot" is empty. Let A_i represent ordered subclasses of left adjuncts of N, i increasing from left to right. Then a C-equivalence is:

(2) A_i ...N C A_i ...N' e.g. blue Venetian glass and Wedgewood pottery

Left adjuncts of N occurring to the left of A_i ...N may adjoin N or both N and N', e.g. in the above example, blue, which is a lower order subclass of adjectives than either Venetian or Wedgewood, might be adjoined to glass or to both glass and pottery, though some speakers may not accept the latter grouping. If the element to the left of A_i...N is an article, it most often adjoins both

---

[1]Cf. Zeno Vendler, "Order of Adjectives."

[2]Possibly a table and chair is T N C (T) N where the second occurrence of T is in zeroed form. This would have to be demonstrated by some test.

N and N', e.g. the Venetian glass and Wedgewood pottery. There would seem to be a restriction on C-equivalence involving N, to the effect that the two entities: N-and-all-its-adjuncts, N'-and-all-its-adjuncts, should be equally specific; this is difficult to state but may be worth investigating. E.g. not early dawn and the twilight, perhaps not dawn and late twilight.

Among the replacement strings of N there are two groups: (1) verb containing strings, like K_4 N V^- (what he said) and (2) non-verb containing strings, namely

R	pronoun	e.g. we are here
Q	some classes as left adjuncts of N	much is known
Q_t		all is forgiven
T_4		either will do
$T_2\bar{A}$		the following is true

A verb-containing replacement string may contain a C-equivalence (what he and the others said); the C-equivalences which apply are those stated for verb-containing strings in general with the one restriction that the scope of the equivalence is determined by the position of the replaced N (see 3.5 and C-equivalences of strings, below).

With reference to the non-verb-containing replacement strings, a C can only occur at the end of the string since the strings are one symbol in length except for $T_2\bar{\bar{A}}$, and T_2 C T_2 does not occur. Let [rn] stand for any replacement string of N; when C occurs at the end of a replacement string we have the C-equivalances:

(3)	N	C	[rn]′	The man and what he stands for
(4)	[rn]	C	N′	We and others of our opinion
(5)	[rn]	C	[rn]′	The following and whatever is based on it

In the first equivalence there is no ambiguity as to the associativity of left adjuncts of N since replacement strings do not have left adjuncts. In the second equivalence, right adjuncts of N (not separated by a comma) are most often associated with N.

3.3 C-Equivalences of the verb phrase

The C-equivalences of verbal elements requires considering the possible object strings of a verb as well as the verb form itself. The verb part can occur in various forms:

V_t	(tense-marked verb)		e.g. goes
W V ∅	(V ∅	verb with no tense mark	
	W	auxiliaries, e.g. can, will, should, may)	e.g. will go
have V-en			e.g. has gone
be V-ing			e.g. is going
do V ∅			e.g. did go

The elements have, be, do themselves occur in one of the above forms. In particular, each may occur in any form written above it in the listing above except do which is always V_t, e.g. will have gone, may have been going (but not are having gone and not have done go). We will name the verb in its various forms wV where for each form of V listed above w has the value of the element preceding V, i.e. $w = \phi$ (before V_t), w = W or do (before $V\phi$); w = have (before V-en); w = be (before V-ing).

We first note several C-equivalences involving W

(1) W C W' e.g. we can and should succeed

Sometimes W is C-equivalence with a verb construction, $V_{\ldots\phi}$, which is followed by $V\phi$. Note: In general, if each member of a C-equivalence requires that a particular element (the same element for each member) occur somewhere to the right of it in the sentence formula, we will show the common required element in { } to the right of the equivalances.

(2) W C $V'_{\ldots\phi}$ $\{V\phi\}$ e.g. They could but are not likely to perform his music

(3) $V_{\ldots\phi}$ C W' $\{V\phi\}$ e.g. We want to and can succeed

Two wV, or two V-ing (G), or two V-en (S), are potentially C-equivalent if among the possible object strings of each V of the pair (as given, say, in a grammatical dictionary) there is a common member. For example, if V_N = a verb which may have N for object, a C-equivalence is:

wV_N C $wV_{N'}$ $\{N\}$ e.g. He produced and directed a film

More generally if γ stands for the subclassification of verbs which can take as object string

(4) wV_γ C $wV_\gamma{}'$ $\{\gamma\}$

If wV_γ is in a string which calls for omission of N from the object string of V, i.e. a string containing V^-, then $V_\gamma{}'$ must also omit an N from its object string, e.g.

The book which we requested and bought.

3.4 C-Equivalences of Strings

We consider a generalized formula for a string β

$$\beta = K\ N_s\ w\ V_{i-}^{+}\ O_i$$

where K = a string identifier and any other elements of the string which precede N_s, N_s = subject N, wV = verb form as defined in 3.3 above, V_i = a particular English verb, O_i = a possible object string of V_i. The verb may call for an object string in reference form (V^+) or for an object string with N-omission (V^-). In particular strings, one or more elements may not be present, e.g. The center string $N_s wV_i^+ O_i$ has no identifier, the who- string $K_1 wV_i^+ O_i$ has no N_s, etc. For these cases, rewrite the equivalences stated below without the general element not present in the particular string and drop duplicates.

Suppose C occurs in the sentence formula immediately following O_i. We have the (type I) C-equivalences:[1]

---

[1] A more general formulation of Type I C-Equivalence for strings is: Let β stand for a string whose elements are from left to right $b_1, b_2 \ldots b_i, \ldots b_n$. Let ${}_i\beta_j$ stand for the elements of β, b_i through b_j inclusive. Suppose in a sentence formula, C occurs immediately following the m^{th} element of β, the type I C-equivalences are:

$${}_i\beta_m\ C\ {}_i\beta_m\ \{{}_{m+1}\beta_n\}$$, i taking on all integer values from 1 through m.

In the example in the text we used the specific string $K_5 NN_s V_i^-\ O_i$. For the same string we show the results of using the more general formulation of type I C-equivalences.

$K_5 NN_s wV_i^-\ O_i$	C	O_i'	see text
"	C	$V_j O_j'$	"
"	C	$wV_j O_j'$	"
"	C	$N_s wV_j O_j'$	"
"	C	$NN_s wV_j O_j'$	whose painting the critics will consider a masterpiece and sculpture the museum will buy.
"	C	$K_5 NN_s wV_j O_n'$	see text

(this footnote is continued on page 31)

$KN_swV_iO_i$	C	$(KN_swV_i)O_i$	whose painting the critics will consider a masterpiece and (whose painting the critics will consider) a discovery
"	C	$(KN_sw)V_jO_j$	whose painting the critics will consider a mastepiece and (whose painting the critics will) acclaim
"	C	$(KN_s)wV_jO_j$	whose painting the critics will consider a masterpiece and (whose painting the critics) will acclaim
"	C	$(K)N_swV_jO_j$	whose painting the critics will consider a masterpiece and (whose painting) the public will reject
"	C	$KN_swV_jO_j$	whose painting the critics will consider a masterpiece and whose courage the public will admire

Suppose C occurs in the sentence formula immediately following V_i; we have:

[In the example we use the string $K_2N_swV_i^- O_i$]

KN_swV_i	C	$(KN_sw)V_i$	$\{O_i\}$	whom the party will nominate and (whom the party will) elect president
"	C	$(KN_s)wV_i$	$\{O_i\}$	" " " " " " (whom the party) may elect president
"	C	$(K)N_swV_i$	$\{O_i\}$	" " " " " " but (whom) people won't elect president
"	C	KN_swV_i	$\{O_i\}$	" " " " " " but whom people won't elect president

Similarly, if C occurs in the sentence formula immediately following w, we have

KN_sw	C	$(KN_s)w$	$\{V_iO_i\}$	whom she can but (whom she) may not invite
"	C	$(K)N_sw$	{ " }	" " " " (whom) he can't help
"	C	KN_sw	{ " }	" " " " whom they won't recommend

The general pattern of type I C-equivalences is clear.[1] For C occurring in the sentence formula somewhere among the elements of string β, what occurs to the right of C in the sentence formula is a structural parallel (repetition of the class marks) of some one of a set of segments derivable from the portion of β which occurs to the left of C. In particular the set contains every different segment of β on the left which reaches up to C. If some member(s) of β have not yet occurred in the sentence formula, i.e. are still required, these elements are expected to follow the occurrence of the C-equivalent elements in the sentence formula. The portion of β on the left which is not structurally paralleled on the right may be said to be present on the right in zeroed form.

Should C occur inside an object string, if the Object String is not of the β-type, the equivalences are those listed for the verb phrase.[2] If the object string is of the β-type, the equivalences which apply are those now being stated (type I) for strings, but in addition the equivalences may be applied to the string of which the object string is a member, e.g. <u>We knew that he, and assumed that she, had read the article</u>. This statement, is equivalent to the "scope" restrictions on object strings and replacement strings stated in 3.5 below. There is sometimes a question as to whether an object string equivalence can be expanded to make an equivalence of two strings on the level of the string containing the object string. E.g. <u>He based</u>

---

[1]See also footnote on p. 30.

[2]See section 3.3 above, formula (4) and the discussion following (4), pp. 22, 3.

<u>his conclusions on their data and my theories</u>. Is this: <u>He based his conclusions on their data</u> and (he based his conclusions on) <u>my theories</u>?

When C occurs in the sentence formula at the end of a string β (i.e. immediately following O_i), a second type of C-equivalence may be found.

<u>Type II C-equivalence</u>[1]

IIa.	$N_s wV_i - O_i$	C $N_s(wV_i \pm O_i)$ e.g.	His first book succeeded and the second one (succeeded) too
IIb.	"	C $N_s w(V_i \pm O_i)$ e.g.	" " " " but the second one didn't (succeed).
IIc.	"	C $N_s(wV_i \pm)O_i$ e.g.	He wrote novels and she (wrote) poems.

Note that in type II equivalences the initial portion of β is paralleled on the right and a terminal or middle portion is zeroed, whereas in type I equivalences, a terminal portion is paralled (sometimes in part shared) on the right and an initial portion is zeroed. In strings like $K_1 wV^+$ which have no N_s the initial portion is the w and only IIb[2] is possible. (e.g. <u>Those who could have helped but didn't</u>).

It is interesting that one formally possible equivalence does not occur, namely

$N_s wV_i \pm O_i$ C $N_s wV_j \pm (O_i)$ e.g. He shattered the glass and she mended (the glass).

---

[1]Type II equivalences are stated for the general form of a string.

$$\beta = N_s wV_i O_i$$

The string identifier K may or may not occur in any of the equivalences.

[2]N_s is dropped from the equivalence. See p. 30.

We also find that certain V in different subclassifications γ, where the γs have a common terminal portion, sometimes (but seemingly rarely) enter into C-equivalences. We give a few examples which range from comfortable to impossible.

wV_N	$C\ wV_{PN}\ P'$	$\{N\}$	e.g. They engaged and relied on assistants
$wV_{PN}\ P$	$C\ wV_N'$	$\{N\}$	e.g. They listened to and recorded conversations
$wV_{to\ V^+}$	$C\ wV_{N\ to\ V^+}\ N'$	$\{to\ V^+\}$	e.g. He likes and has the desire to work
$wV_{N\ to\ V^+}\ N$	$C\ wV_{to\ V}'$	$\{to\ V^+\}$	e.g. Having the desire and meaning to work
wV_{PN}	$C\ wV_{NPN}\ N'$	$\{PN\}$	e.g. They relied and based their conclusions on testimony from paid witnesses
$wV_{NPN}\ N$	$C\ wV_{PN}'$	$\{PN\}$	e.g. They referred decisions and deferred to their senior
V_A	$C\ wV_{NA}\ N'$	$\{A\}$	e.g. They seemed and considered others foolish
wV_{NN}'	$C\ wV_{NN}'$[1]	$\{NN\}$	e.g. He gave and considered her a pearl
wV_N	$C\ wV_N'$[1]	$\{NN\}$	e.g. He saw and seemed a native

We sometimes find a C-equivalence involving an adjunct PN and an object string which ends in N, e.g.

They bought tickets for and went to see Marcel Morceau
$P \quad C \quad V \quad to \quad V_N \quad \{N\}$

The boys who liked and shined the shoes of soldiers
$V_N \quad C \quad V_N \quad N \quad P \quad \{N\}$

Left adjuncts of V which occur to the left of V or w in a C-equivalence may adjoin V or both V and V':

$wV \quad C\ wV'$ e.g. he will quickly return and report to us

e.g. they often watch or listen to the rehearsal

---

[1]There are two different verb-subclasses having the form N, and two having the form NN.

Why? Possibly what is involved is the fact that a verb has not one possible object string but a set of possible object strings. In general, when a symbol (class) occurs on both sides of C, it is not the same word member which occurs on both sides, but different word members (see examples). So in the equivalence being discussed: V_i occurs on the left and V_j on the right, both are in the same verb class V_γ which has γ as object string, i.e. appears on the list of object strings for both V_i and V_j. O_i is an instance of γ. The string grammar allows a verb to occur with any 1 of its object strings. Therefore, as soon as V_j occurs in the sentence formula, as far as the grammar is concerned, it is free to choose any one of its objects. But the equivalence relies on O_i being the chosen object of V_j. There is nothing in the string grammar to assure this will be the case, hence the equivalence is not usable.[4]

A few examples should indicate how type II equivalences look in adjunct strings.

Type of C-Equivalence	String	Example
II a	$K_5NN_swV_i{}^{-}O_i$	A teacher whose methods the children like and the mothers also
II b	$K_3N_swV_i{}^{-}O_i$	The route which the guide recommended but you didn't.
II c	$pK_2N_swV_i{}^{+}O_i$	A man with whom she discusses politics and he art

---

[4]There are some reasons for considering <u>be</u> to be an incomplete verb, rather like w. In line with this we find the C-equivalence

$$N_swV_i \pm O_i \qquad C \qquad N_swV_i \pm (O_i)$$

for the verb <u>be</u> but no other verb. E.g. we have

<u>He was a poet but she wasn't</u> (a poet).

but not

<u>He wrote poetry but she didn't write</u> (poetry).

$K\,N_s$ to V_i-O_i	C	O_1'	whose painting the critics will consider a masterpiece and a discovery
"	C	V_j^--O_j'	" " " " " " " " and acclaim
"	C	wV_j^--O_j'	" " " " " " " " and will acclaim
"	C	$N_s wV_j^-$-O_j'	" " " " " " " " and the people will reject
"	C	$KN_s wV_j^-$-O_j'	" " " " " " " " and whose courage the public will admire

The above C-equivalence can be shown to consist of two whole strings conjoined by C in which those elements on the left which do not appear on the right are actually there in zeroed form. The zeroed elements are not only the same structural entities but the same words as are on the left in the analogous position. We repeat the above C-equivalences showing the zeroed elements in parentheses.

E.g. for m = 3, $\beta = K_5 NN_s V_i^- O_i$, we would have

i = 3	$K_5 NN_s$	C	N_s	$\{wV_i^-\ O_i\}$ whose painting the critics but not the public admired
i = 2	$K_5 NN_s$	C	NN_s	{ " } " " " " and sculpture the public admired
i = 1	$K_5 NN_s$	C	$K_5 NN_s$	{ " } " " " " and whose courage the public admired

Note that the equivalence for i=2 is admitted by the general formulation but not by the formulation in the text.

(3)	$N_s wV_i \; N_s wV_j O_j$	C	$N_s(wV_i) \; N_s(wV_j)O_j$	e.g. We hoped the green team would beat the red and you (hoped) the red(team would beat) the green.
(4)	$N_s wV_N N \; K_1 wV_j O_j$	C	$N_s(wV_N)N \; K_1 w \; V_j O_j$	e.g. We interviewed people who witnessed the events and you (interviewed) people who didn't (witness the events).

In each of the above C-equivalences, on the left there are two strings, one including the other, and on the right a similar construction, with a type II C-equivalence holding between each pair of strings in corresponding positions.

Double C-equivalence	Inclusion Relation	Including-string C-Equivalence	Included-string C-equivalence	Notes
(1)	Second string is object string of verb in first string	Hc	IIa	
(2)	"	"	IIb	<u>That</u> (sentence introducer of object string) must be zeroed
(3)	"	"	IIc	selection of contrasting word pairs needed
(4)	Second string is adjunct of element in the first string	"	IIb	More comfortable than IIa or IIc in included string

Notice that (1)--(3) exhaust the possibilities for the object inclusion relation since IIa and IIb applied to the including string would result in losing the included string. Theoretically adjunct inclusion should permit all combinations of type II equivalences but actually few occur.

In addition to the equivalences discussed so far, there is one "limiting case" of zeroing. For a particular class of sentence adjuncts $\overline{D}$ D, PN, PA, to V^-, and strings beginning with $\not{c}_2$ or $\not{c}_3$ (see table____), we may have I'.

$N_s w V_i O_i$ C $(N_s\ w\ V_i\ O_i)\ \overline{D}$ e.g. We left and (we left) fast

e.g. We bought them ice cream but (we bought them ice cream) to keep them quiet.

When C occurs at the end of a string, we have a further (type III) C-equivalence. Let $[\beta]$ stand for any member of a given class of strings, e.g. right adjuncts of N, sentence adjuncts. Then

III $[\beta]$ C $[\beta]$ e.g. The delegates elected but not receiving an absolute majority (strings S^-, G^-)

Actually not all $[\beta]$ are C-equivalent. We find more awkward: <u>The delegates elected but who did not receive an absolute majority</u> (strings S^+, K_1V^+) and almost impossible: <u>The man whose car we borrowed and from Philadelphia</u> (Strings K_5NNV^-, PN).

<u>Double zeroing</u>

Consider the following "double C-equivalences" (zeroed elements in parentheses):

(1) $N_s w V_i [N_s w V_j O_j]$ C $N_s (w V_i) [N_s (w V_j O_j)]$ e.g. I claim that the police were carrying arms and you (claim that) the demonstrators (were carrying arms).

(2) $N_s w V_i [N_s w V_j O_j]$ C $N_s (w V_i) [N_s w (V_j O_j)\ ]$ e.g. I claim that the men carried a flag and you (claim) that they didn't (carry a flag).

3.5 Scope of C-Equivalences

If we consider α a variable name for the structure to the left of C in the sentence formula which is paralleled by a structure (α') to the right of C, then the range of the values of α can be stated in terms of a string analysis of the sentence formula to the left of C. (This is convenient for a left to right analysis).

Let β_i designate a string which is represented (= has at least one member) to the left of C in the sentence formula, i increasing leftward from C, i=o, 1, 2... . Further, let k stand for the relative depth of a string, i.e. if β_i is at depth k and string β_j is an adjunct of β_i or of a symbol in β_i, then β_j is at depth k + i.

Given a string structural analysis of a sentence formula to the left of C, the determination of α in the C-equivalence α C α' is as follows:

1. If C falls inside β_o (= not at the end of β_o), then α is in β_o and the C-equivalence which applies is type I.

2. If C falls at the end of β_o, and β_o is an adjunct string, then α lies in some β_i whose relative depth k_i does not exceed the relative depth k_j of a β_j represented between β_i and C; the C-equivalences which apply are types I, II, III.

3. If C falls at the end of β_o, and β_o is an object string or a replacement string, we can determine where α lies, using 1 and 2 above as follows:

   a) Apply 2, as if there is no i > o. This covers the possibility $\alpha \subset \beta_o$.

   b) Replace β_o by one symbol (O if β_o = object, N if β_o = replacement of N). C now falls inside or at the end of β_1, the string which includes O or N. Substitute β_1 for β_o in 1. and 2. above, and apply 1 and 2.

<u>Example</u>. We show a set of sentences which have the same initial portion up to C but differ in the portions following C because of different values of α, The common initial portion is shown analyzed as to its string structure, preceding the set of sentences. Relative depth is shown by writing a string of depth K + 1 below a string of depth k; strings represented to the left of C are numbered according to the convention stated at the beginning of this section.

<u>Initial segment</u>:

$\overset{N}{\text{A group}}$ β_2 $\qquad\qquad\qquad$ $\overset{V^{+}}{\text{performed an experiment}}$ β_2

$\qquad$ $\overset{P\quad N}{\text{of chemists}}$ β_4 $\qquad\qquad\qquad$ $\overset{K\quad N\quad V^{-}}{\text{which we repeated}}$ β_1

$\qquad\qquad$ $\overset{P\quad N}{\text{in Washington}}$ β_3 $\qquad\qquad\qquad$ $\overset{D}{\text{yesterday}}$ β_o

Table of Example Sentences

1 String which contains α	2 Type of C-equivalence	3 C-Equivalence α C α'	4 Sentence illustrating the C-Equivalence of column 3
$\beta_o = D$	I	D C D'	A group of chemists in Washington performed an experiment which we repeated yesterday and today
$\beta_1 = K_3N_swV^{-}_iO_i$	I	$K_3N_swV^{-}_iO_i$ C $wV^{-}_jO'_j$	A group of chemists in Washington performed an experiment which we repeated yesterday and are now analyzing
	I	$K_3N_swV^{-}_iO_i$ C $N_swV^{-}_jO'_j$	A group of chemists in Washington performed an experiment which we repeated yesterday and other laboratories have also repeated
	I	$K_3N_swV^{-}_iO_i$ C $K_3N_swV^{-}_jO'_j$	A group of chemists in Washington performed an experiment which we repeated yesterday and which others may find interesting
	I'	$K_3N_swV^{-}_iO_i$ C $\bar{D}$	A group of chemists in Washington performed an experiment which we repeated yesterday and with result
	IIa	$K_3N_swV^{-}_iO_i$ C N'_s	A group of chemists in Washington performed an experiment which we repeated yesterday and others also
	IIb	$K_3N_swV^{-}_iO_i$ C N_sw'	A group of chemists in Washington performed an experiment which we repeated yesterday and others will also
	IIc	$K_3N_swV^{-}_iO_i$ C N_sO_i doesn't apply since O_i is zero	
	III	[r. adj. of N] C [r. adj. of N]'	A group of chemists in Washington performed an experiment which we repeated yesterday and whose results may interest others

Table of Example Sentences (continued)

1 String which contains α	2 Type of C-Equivalence	3 C-Equivalence α C α	4 Sentence illustrating the C-Equivalence of column 3
β_2 $B_s wV_i O_i$	I	$N_s wV_i O_i$ C O'_i	A group of chemists in Washington performed an experiment which we repeated yesterday and certain calculations
	I	$N_s wV_i O_i$ C $V_j O'_j$ same as following line in this case	
	I	$N_s wV_i O_i$ C $wV_j O_j$	A group of chemists in Washington performed an experiment which we repeated yesterday and made certain calculations
	I	$N_s wV_i O_i$ C $wV_j O_j$	A group of chemists in Washington performed an experiment which we repeated yesterday and they will be pleased to know the results checked
	I'	$N_s wV_i O_i$ C $\bar{D}$	A group of chemists in Washington performed an experiment which we repeated yesterday and carefully
	IIa	$N_s wV_i O_i$ C N_s	A group of chemists in Washington performed an experiment which we repeated yesterday and another group also
	IIb	$N_s wV_i O_i$ C $N_s w$	A group of chemists in Washington performed an experiment which we repeated yesterday and another group did also
	IIc	$N_s wV_i O_i$ C $N_s O_i$	A group of chemists in Washington performed an experiment which we repeated yesterday and a group of mathematicians the calculations

For $\beta = 2$, the equivalences with zeroings are quite awkward. One might say that as the two parts of a C-equivalence become separated, the occurrence of the C-equivalence with zeroings become less likely.

Notice that for β_3 and β_4 no C-equivalence is possible because the depth of β_3 and β_4 are each greater than the depth of a string to their right, β_2. E.g. we could not have

β_3 A group of chemists in Washington performed an experiment which we repeated yesterday and with whom we have corresponded.

β_4 A group of chemists in Washington performed an experiment which we repeated yesterday and of biologists.

3.6 Scope markers

Sometimes the scope of a C-equivalence is marked in the sentence formula by the occurrence of a particular symbol; then the scope rules of 3.5 apply only to the stretch of the sentence formula between C and the scope marker.

The idea of scope markers is presented here as a hypothesis to explain the function of certain words which sometimes occur in disjoint association with particular conjunctions. These words also occur in sentence where they are not in special relation to conjunctions, but it seems that the two functions can be separated; either the word occurs as part of a nounphrase, or it occurs in a wide distribution where it has no grammatical function other than marking the scope of a C-equivalence.

In this section three pairs of conjunction-and-scope-marker are considered

Scope Marker	Associated with	Example of Scope-marked C-equivalences
both	and	They both plastered and painted the walls
either	or	Either he or I will go
neither	nor	Neither God nor country saved him

The conjunction but has no scope-marker. It may turn out that the pairs er...than[1] and as...as (comparatives can be treated as conjunction plus scope-marker. On

[1]The comparative morpheme -er occurs in A-er (prettier, finer), more, less. rather, worse, D-er, and possibly other forms.

these we first need to know in some detail how the C-equivalences and morphophonemic zeroing operate.

The conjunctions _and_, _or_ can each occur without a preceding scope marker, e.g. _both he and I_, _he and I_; however, if _both_ or _either_ or _neither_ occurs in a sentence, it must be followed, at some distance, by its associated conjunction. The case with _er...than_ is the opposite: the conjunction _than_ does not occur except preceded by _-er_; the _-er_ morpheme, however, may occur without a following _than_, e.g. _This is more likely the case_.

When a scope marker is followed by two occurrences of its associated C in a sentence formula, e.g. _Either he or I will buy or borrow a pen or pencil_, there is a question as to whether the scope marker functions in association with one C or with more than one C. A starting guess is that it operates independently with each C.

As we noted, the words _either_, _neither_, and _both_ have more than one classification, e.g. in _Either type will do_, _either_ is an article, T, and in _Either he will go or I will_, it is a scope marker (henceforth written Σ). A decision can sometimes be made because of the environment. The following is a preliminary list of environments and classifications.[1]

Word	occurs	classification	example
either, _neither_	Before T	Σ	_Either the matter will be settled soon or not at all_
	Before $\cancel{R}$ (pronoun)	Σ	_Either he or I will go_
	Before N_{pl} (plural noun)	Σ	_Either children are disciplined or spoiled_
	Before N_1 (count noun)	T_4[2]	_Either type will do_
	Before $N_{sg', \neq 1}$ (singular, not count noun)	T_4/Σ	_Either language can be used._ _Either language is described formally or culturally_
both	Before T	Q_T	_Both the men are literate_
	Before _of_	Q_T	_Both of the men are literate_
	Before $\bar{R}$	Σ	_Both we and they attended_
	Before N_{pl}	Q_T/Σ	_Both men came_ _Both men and women came_

[1] Where N appears is an environment, it is to be understood that $\overline{AN} = N$ and that N is not a left portion of a compound noun.

[2] _Either_ and _neither_ also occur as a noun (_either will do_). In TDAP 27 T_4 (_either_, _neither_) is a subclass of T which also is listed among the replacement strings of N. Similarly Q_T may occur as a noun. See p. 21, p. 26.

4. Left Adjuncts and the Analysis of the Nounphrase

For present purposes the left adjuncts of N can be divided into 3 groups.

a) Article-quantifier classes T, B, Q_T, D_T, Q.

b) Adjectival elements: A, S, G, Z, N's

c) The left portion of compound nouns: N, NA, NS, NG or any combination of these

Each of these presents certain special features, or problems for string computation. For a) it is a question just now of obtaining a more refined classification of the elements and establishing their relative order of occurrence in the nounphrase. In b) there are sometimes ambiguities of an associational kind, e.g. in hazy blue sky, is the sky hazy or is the blue hazy? The problem here is to state for which elements in which positions more than 1 grouping of the elements is possible and to provide a systematic procedure for obtaining all the groupings (and perhaps to select the more probable among the groupings). In c) there are two questions, one which concerns the analysis of the compound noun itself (more a transformational than a string analysis problem) and another which involves ambiguous sequences, e.g. does a particular sequence of Ns constitute one nounphrase or more than one nounphrase, for example flint tools is a compound noun in the flint tools were made into spear-like shapes, but is two nounphrases in the flint tools were made of was of local origin. Of these problems, those concerning b) and c) are discussed below.

4.1 Adjectival elements

Sequences of A, S, G elements in a nounphrase are in general independent of each other and adjoin an element farther to the right, i.e. A, S, G, elements are not nested among themselves.[1] Thus there is only one grouping for the

---

[1] A sequence of pure As in a noun phrase sequence presents no problem in grouping; each A is an independent adjunct either of the first N's to the right or of the "head" (rightmost) noun of the noun phrase. However, when the verb-derived adjectival elements S, G are introduced into the sequence, there may be some question as to the grouping. Most often these, too, are independent,

```
e.g.  The measured optical density
               S        A       N
          (        ) (     )

      additional reverberating circuits
           A              G          N
      (        )    (          )
```

But we find

```
slow moving plane     the assumed correct results     long-lasting facilitation
  A    G      N                 S       A      N           A    G         N
((  )       )         ((              )       )        ((  )            )
                      (               ) (     )
```

where the elements are not grouped independently.

(footnote continued on page 45)

A, S, G in each of the following examples:

```
cheerful young man
   A      A    N
(      ) (   )

The outlying newly constructed residential districts.
       G        D         S          A        N
    (       ) ((  )          ) (        )
```

However, certain adjectival elements may themselves have left adjectival adjuncts, e.g. Z(color words), so that if a sequence of A, S, G elements is followed by Z N, several different groupings are possible: the A S G sequence (or certain parts of it) may adjoin Z or may adjoin N. Making a separate class Z is in part based on this distributional fact.[2]

---

1 (continued from page 44)Transformational analysis shows that slow moving plane is a transform of a single sentence plane moves slowly whereas the measured optical density is transformationally derived from two sentences: The density is optical and the density is measured. Similarly compound nouns involving verb-derived elements are often transforms of a single sentence which contains all three words, e.g. hair splitting distinction from distinction splits hairs. It is a question whether string analyses can distinguish among types of compound nouns and types of adjectival sequences.

2 The distributional characteristics of color word Z which distinguish them from pure nouns like hat, mass and from pure adjectives like young, cheerful are:

They can occur as an adjective, e.g. blue dress
They can occur as a noun, e.g. Blue is my favorite color
They can draw left adjuncts from among A, S, G, Z (in a noun phrase)
e.g. pale blue dress
e.g. frosted blue color
startling bluelake
green blue sky
They can occur as the left portion of a compound noun, e.g. blue period
[The latter is distinguished from the adjective occurrence by stress patterns. For instance, contrast the painter's blue mood with Picasso's blue period].

On the same grounds another class is suggested, call it Y, of words like stone, copper (roughly, words describing materials). They occur, corresponding to the above,

As adjective e.g. stone wall
As noun e.g. Stone is stronger than brick
With left adjuncts in the Nounphrase
e.g. hard stone wall
chipped stone tools
flaking stone walls
tungsten steel plate
As the left portion of a compound noun, e.g. stone age
[To distinguish the latter from the adjective occurrences, contrast stone wall with stone age, copper tubing with copper compound].

The following example shows the possible groupings of A, S, G elements preceding Z N and how the groupings increase as a function of the number of A, S, G elements. If there are n A, S, G elements preceding Z N, there are n+1 groupings of the entire sequence.

```
                    The      hazy      blue      sky
                            (    )    (    )
                           ((    )         )

          The      pale     hazy     blue      sky
                  (    )   (    )   (    )
                  (    )  ((    )        )
                 ((    )   (    )        )

     The     fading      pale      hazy     blue      sky
            (      )    (    )    (    )   (    )
            (      )    (    )   ((    )        )
            (      )   ((    )    (    )        )
           ((      )    (    )    (    )        )
```

The element N's may also have left adjuncts, but unlike Z, any A, S, G which occurs to the left of 1 or more N's, adjoins the left-most N's, and may not equally well adjoin the head N of the nounphrase. Hence for n A, S, G elements preceding N's, N, there is one grouping

```
        e.g.  Underdeveloped country's youth
             ((              )         )
```

One might accept a second grouping of the above example

```
        e.g.  underdeveloped country's youth
             (              ) (        )
```

in which underdeveloped is associated with youth, but I believe such an intended association would occur as country's undeveloped youth.

One notes that the groupings of A S G elements before Z N is the complement of the groupings of A S G elements before N's N, with regard to the set of all possible groupings which do not violate non-intercalation. If this has significance, perhaps N's and Z represent two different ways of allowing adjectival elements to adjoin other adjectival elements.

c) Compound noun

We now consider sequences of noun phrase symbols which might contain compound nouns, i.e. which have on the right NN or NAN or NGN or NSN or combinations of any of these. There is a question as to whether such a sequence constitutes one or more noun phrases. We discuss below several restrictions which can be made in making the decision.

1. Count noun restriction: the sequence cannot be broken in such a way as to leave a count noun in the final (rightmost) noun position with no T or B preceding it in the noun phrase.

Frequent cases:

a) The final (rightmost) noun of the sequence is a count noun. Therefore, the whole sequence between article (T or B) and final noun is one noun phrase. Example of sequences decided by (a) are:

A 10-40 percent w/r sucrose density gradient was prepared according to the method of B. & L.

What he thought as a syndicalist in the USA he did as a trade union organizer in Ireland.

Connolly never knew the vanguard party.

The paintings and engravings at Lascaux are attributed to the peak period of Perogordian art.

Some blackist marks which resembled the deposit of a grease lamp were found.

The sedimentation coefficient of mosaic virus was measured in a Spinco model E analytical ultra centrifuge.

b) There is no initial T or B; therefore the sequence cannot be broken in a way that leaves the left portion terminating in a count noun.

Examples of (b) are:

The date of Lascaux within the framework of Franco-Cantabrian wall art can be estimated by a study of the styles and superpositions of the paintings in relation to the known stages of cave art. (count nouns: wall, cave)

These experiments consisted of infectively and particle assays of fractions obtained by density gradient centrifugation. (count nouns: particle, gradient)

2. Plural: A compound noun does not contain a plural noun in non-final position. Example:

> In the caves remains belonging to Homo Sapiens cover those of the last Neaderthals.

3. Pronouns: A compound noun does not contain pronouns. Example:

   They worked flint blades compared with stone worked flint blades

4. Comma: A compound noun does not contain a comma.

Suppose a sequence can be divided into two (or more) noun phrases on internal grounds. Are further restrictions available by considering the immediate context? We first take up in what cases we find two adjacent noun phrases in a sentence.

Two adjacent noun phrases can occur in the following cases:

a. In a single string

After a verb which has the string NN as one of its possible objects

verbs like give	e.g. They gave the boy books for his birthday
" " elect, make	we elected a woman president
" "	we made the children toys
" " consider	we consider him a fool

In the sentence adjunct NN

e.g. The first book a success, they contracted for a second.

b. At the border of two strings

b1. The second noun phrase is the beginning of a right adjunct of N or of a sentence adjunct.

The right adjuncts of N which begin with N are:

N (apposition)	e.g. the local lesion host chenopodium hybridum the evil man can inflict
NV^-	the evil man can inflict ...
NV^+(only after N_8)	the reason man exists ...
NPK strings	... samples of which we tested

The sentence adjuncts which begin with N are:

NO_B strings (6 in number. See Sect 1. part d.)	e.g. His mission accomplished, he ...
N V_{14}^- (V_{14} = verbs like know)	These results we relaize, are ...
N_7 (N_7 = time words)	People today

Since right adjuncts of N are always permitted after N, and sentence adjuncts are permitted after N in most positions, the occurrence of contiguous Ns in the sentence formula would seem to give rise to many different analyses. Actually most of the above string rarely occur without a comma to set them off from the preceding noun phrase, and some of them occur infrequently even with a comma. For the moment we are only considering sequences of noun phrase elements which could constitute either a single noun phrase or several noun phrases, not the entire problem of consecutive N's, E. g. we consider the evil man can inflict but not the evil a man can inflict, although in the latter,while there is no ambiguity as to the number of noun phrases, the problem of assigning string status to the two noun phrases still exists. The remarks to follow concerning commas and other restrictions apply to sequences of noun phrases in general, but they are stronger for the ambiguous case because often an unambiguous grammatical separation of two noun phrases functions to establish a string boundary whereas when there is no grammatical separation, a comma is practically mandatory.

Certain restrictions on right and sentence adjuncts which begin with N are:

N O_B strings — absolutely rare

almost never occur without a comma separating its initial or final element from adjacent sentence elements.

N apposition — rare without a comma

When the N apposition is a name it sometimes occurs without comma,

e.g. The painter Picasso

The local lesion host chenopodium hybridum

N P K strings — rarely withou a comma

e.g. Bacteriophage γ x 174 top and bottom components the particle densities of which in rubidium chloride are 1.32 and 1.40 respectively, were added to the virus solution.

The particle, the characteristics of which we have described, is indeed the infectuous virus.

N V — never with a comma preceding N

e.g. not: The evil, man can inflict, is ...

The V^- of N V^- implies further restrictions. For example, the evil man exists could not be interpreted to contain an N V^-string because V^- implies the omission of an N from the verb object and exists is intransitive. If the verb following an N N (ambiguous sequence is passive, the NN is almost surely not a noun followed by N V^-because then the verb object would have to be such as to allow two N omissions.

Examples of decisions against N V^-:

Diffusion experiments were carried out.

Apparent diffusion co-efficients were calculated

... in which the nucleoprotein content is determined by ...

Some flint tools were found lying just under the surface of the clay.

b2. The second noun phrase is the beginning of a replacement string of N (N G^+ or N G).

b3. All the other border cases. Actually, since b1 and b2 exhaust the list of strings which begin with N (except N V^+ center string) the only other broader possibility is that the first noun phrase be the end of a string, and the second noun phrase be "the next required member" of the including string. Although a comma is often present, especially if the adjunct ending in N is long, there seems to be no rule to apply. But since cases b1. does have restrictions, if one can establish that the noun phrase sequence in question is not at the end of a string, then the problem of deciding whether an ambiguous sequence constitutes one noun phrase or more than one, is considerably simpler. For example, a noun phrase, sequence at the beginning of a sentence as such a case. It can only contain a string border when the first noun phrase is itself a complete string, such as N_7 (time words)

e.g. Today science is heavily administered,

or certain "expressions"

e.g. no wonder observers have commented that ...

The most frequent uncertain construction where the first N is at the end of a string and the second N is the next member of the including string seems to be P N N: in some cases P N N $\rightarrow$ P NP NP

e.g. The citizen Army was in part members of the Irish Transport Workers Union.

In order to minimize the stress during reversal subjects slept all night on the last control period.

A comprehension of the evil man can inflict is necessary

and in other cases P N N $\rightarrow$ P NP

e.g. From the sedimentation and diffusion coefficients at a virus concentration of 0.42 per cent the molecular weight was calculated.

We conclude on the basis of velocity and equilibrium sedimentation experiments that the particle, the characteristics of which we have described, is indeed the infectuous virus.