

III

THE RELATION BETWEEN
THE THEORY AND ITS
APPLICATION
TO PROGRAMMING SYNTACTIC
ANALYSIS WITH A
DIGITAL COMPUTER

Edited by K. SPARCK JONES

See

The Application of the Cambridge Language Research Unit
Model of Syntactic Description
Edited by K. Sparck Jones
Cambridge Language Research Unit, M.L. 145

CAMBRIDGE LANGUAGE RESEARCH UNIT
20 MILLINGTON ROAD
CAMBRIDGE, ENGLAND (C) 1961

ABSTRACT PART III

In this part the general theoretical model is developed, in terms of constituent types and participation classes, to give methods for constructing syntactic dictionary entries for particular languages.

PART III. THE RELATION BETWEEN THE THEORY AND ITS APPLICATION

TO PROGRAMING SYNTACTIC ANALYSIS WITH A DIGITAL COMPUTER

Substituent types and participation classesFunctions

We concluded the theoretical description of the Cambridge Language Research Unit model for syntactic analysis (1) by giving the full product lattice for classifying substituents, and the algorithms used within it for determining the functions of groups of constituents. We must now show how the coding used in the empirical procedures currently being tested on the Cambridge computer is derived from the theory. The actual program is fully discussed elsewhere (2), but it will be illustrated by a rule-of-thumb example.

The full product lattice has 144 elements, and in principle substituents with any pair of these functions may form a group, or compound substituent. As the model does not take the order of the components of a group into account we thus have $\frac{1}{2}(144^2)$ possible different groups. The function of a group is, however, determined by the meet algorithm, and the group formed by any one of the 10368 possible pairs of substituents must therefore be defined by a point in the lattice. There are thus only 144 different meet-points. 12 of these points, moreover, are in the lower ideal of Z.Z. and are not accepted as they stand but converted by the polar algorithm. This leaves us, therefore, with 132 result-points representing different kinds of group.

By using this set of kinds of group we can extend our classification system. In setting up the syntax lattice we treated it as a schema for classifying

substituents according to their functions. The function of a substituent is naturally related to its behaviour in groups of substituents, but we did not initially attempt to classify substituents according to the kinds of group in which they can figure. It will now be clear that we can give more information about a substituent if we take what we may describe as its grouping possibilities into account. These are derived from the lattice in a straightforward way: for a substituent with a particular function we list the set of result-points which can be reached when operating on a pair of substituents, one of which is the substituent in question. We thus give, in terms of their functions, the kinds of group in which the substituent can participate. This information is represented by a positive mark in the appropriate positions in a 132-place entry. The record is further refined by entering, for each kind of group in which the substituent can figure, whether it functions as governor, or dependent, or either.

For practical purposes, however, the size of entry just described is not very satisfactory: thus it is too large for convenient machine handling, or for teaching to dictionary makers. It can, however, be reduced as follows. In terms of the theory the set of 132 result-points can be naturally divided into those which lie in the principal exponent of the primary lattice, and those which fall elsewhere, that is, into those with secondary function Z and those with neither function Z. (Except for Z.Z. points with primary function Z are excluded by the polar algorithm). It will be clear from the lattice that there are 12 points of the first kind and 120 of the second. This distinction represents the extent to which further grouping is required before the stop-point defined by Z.Z. can be reached. Substituents with secondary function Z can be 'direct' components of full clauses; those with neither function Z require at least one Intermediate grouping, with the application of the polar algorithm, before they can be grouped to

III. 3

give a full clause. It can be argued that the information about a substituent represented by the fact that it can be a member of a group of the second kind is less useful than that representing its membership of a group of the first kind, given that from a group of the second kind one of the first kind will be reached. If we accept this argument, we can then replace the 132-place classification by one with 12 places. For a particular substituent this replacement will give the result-points in the principal exponent of the primary lattice which will be reached by operations on pairs of substituents of which the substituent in question is a member.

The argument just given for reducing the number of kinds of group cannot however be accepted just as it stands: for although the 12 kinds when selected can be distinctively characterised in terms of the theory, the reason for selecting them is not derived from the theory. The fact that the information about a substituent represented by its occurrence in a group with a function defined by a point in the upper ideal of Z.Z. is useful depends on the empirical fact, that, when the structure of sentences as a whole is considered, groups with functions defined by points outside the upper ideal of Z.Z. are in most languages unimportant.

This assertion can, however, be given a theoretical justification, and one which can straightforwardly be described in terms of the lattice model. This theoretical hypothesis is that the number of possible constructions in a language is limited by what we will call the principle depth of avoidance. This principle is discussed by Yngve (3) (he does not actually give it a name): he asserts that at any point in the course of uttering a sentence the minimum number of steps by which a sentence (though not necessarily the one intended) could be completed, must not exceed a certain limit. In terms of bracketting (see below), we can say that each 'step'

represents the closing of a bracket group.

In terms of our lattice model, Yngve's hypothesis means that at any point in the uttering of a sentence we must be within a limited number of steps from Z.Z. If we restrict ourselves to endocentric constructions we cannot in fact be more than 3 steps away, as this is the maximum number of steps possible between Z.Z. and the top of the lattice. With exocentric constructions, on the other hand, we can in principle be much further from Z.Z. At the word "very" in the sentence "people living in very large houses must be rich", for instance, we are 6 steps from Z.Z. This is shown by the bracket structure as follows:

(people (living (in ((very large } houses))) } ((must be) rich)

The number of steps required to complete the sentence increases with the depth, i.e. with the complication of structure; for the more complicated the structure the greater the restrictions on the related constructions, and thus on steps by which the sentence can be completed. In contrast, if the depth decreases the constructions relax). Yngve's hypothesis is that in general we try to avoid this kind of situation.

In the syntax lattice, if we are outside the ideals of Z.Z. at least 1 step will be required to reach them; points in the upper ideal of IC. IC are in fact at least 2 steps from them. Applying Yngve's theory to the lattice, therefore, we can say that the language is biased against constructions with functions defined by points outside the ideals of Z.Z. We then have a theoretical justification, rather than an empirical one, for the assertion that functions defined by points in the upper ideal of Z.Z. are sufficient for classifying compound substituents.

Constitutions

The classification of groups only by their function is nevertheless too crude; we must also take their constitution into account. The simplest way of doing this would be to allow two alternatives for each kind of group, endocentric and exocentric. This by itself is in practice inadequate, and a more refined system of distinctions is required. In our previous discussions we made use of different 'areas' of the lattice; now one important area or aspect of a product lattice is its centre: the set of points representing the bounds and the central elements. In our 2-factor product there are 4 of these, the bounds IC.Z and Z.IC and the centrals IC.IC and Z.Z. * (It should be noted that the centre elements themselves constitute a Boolean lattice). We can say that the centre of a product lattice contains the most extreme contrasts given by the system as a whole; thus, in our case, Z.Z, the function defining the full clause, marks the most complete and independent syntactic group; IC.IC, the conjunction function, marks the least independent group. Now given this set of functions, with their characteristic properties, it is clear that we can obtain useful information about the nature of a group if we relate it to the centre: i.e. we can ask whether any of its components are members of the centre; more specifically, we can ask whether both components, or either or neither, are in the centre, and if they are, which central elements they fall on. This will give us the following classification:

9 Constitutions

a) both components in the centre

COMPOUND CONSTRUCTION

* The terms used here to distinguish different elements in the centre are not very satisfactory, but they are taken over from the theoretical paper.

III. 6

- | | | |
|-------------------------------------|-------|-----------------------|
| b) one component in the centre - on | IC.Z | CONJUNCT CONSTRUCTION |
| " " " " - on | Z.Z | SUBJUNCT " |
| " " " " " - on | Z.IC | } these do not occur |
| " " " " " - on | IC.IC | |
- c) neither component in the centre
- | | | |
|-------------------------|---------|------------|
| i) exocentric functions | S.Z,O.Z | Z - CLAUSE |
| " " | S.X,O.Z | S - CLAUSE |
| " " | S.Z,O.X | O - CLAUSE |
| " " | S.X,O.X | I - CLAUSE |
- where X is any function other than Z
- | | |
|-----------------------|--------|
| ii) endocentric | GROUP |
| d) one component only | SIMPLE |

One point should be noticed: the endocentric-exocentric division is only made where groups with neither component in the centre are concerned; for here the information derived from membership of the centre *is* only negative, and some other Information must be given.

The interrelation of function and constitution to give substituent types

We must now relate function and constitution. We have 12 kinds of group and 9 constitutions, and in theory each kind of group can have each kind of constitution. This would give us 108 possible alternatives, or substituent types. Closer inspection shows, however, that these are not all genuine possibilities: thus a Z-CLAUSE has components with functions S.Z and O.Z; since the meet of these is Z.Z, this is the only function a Z-CLAUSE can have. Conversely, S-, O- and I-CLAUSES cannot have the function Z.Z.

When we have eliminated these we are left with a much smaller number of

alternatives. Empirical facts about actual languages can now be brought to bear: as far as we know at present only about 20 of these alternatives actually occur; and it has been found that only 12 of these are required for English. We are thus left with a comparatively small set of substituent types for making syntactic dictionary entries. The entry for a particular substituent is called its participation class. In English, therefore, we will have 12-place participation class entries.

The 12 substituent types used for English in the current experiments are as follows:

1.	Conjunct Groups	:	these are a special case as they take the character of their constituents.
2.	Adverbial Groups	:	function IB.Z constitution GROUP
3.	Adverbial Clause	:	" IB.Z " O-CLAUSE
4.	Adjunct Group	:	" SA.Z " GROUP
5.	Participial Clause	:	" IA.Z " O-CLAUSE
6.	Relative Clause	:	" SA.Z " S-CLAUSE
7.	Nominal Group	:	" S.Z " GROUP
8.	Infinitive Clause	:	" S.Z " O-CLAUSE
9.	Verbal Group	:	" O.Z " GROUP
10.	Predicate	:	" O.Z " S-CLAUSE
11.	Marked Clause	:	" IB.Z&S.Z " SUBJUNCT
12.	Free Clause	:	" Z.Z " Z-CLAUSE

English substituent types

The 12 substituent types required for English are as follows:

1. Conjunct Groups

These are groups in which two or more words or phrases of the same grammatical function are joined by a conjunction.

E.G. "you and he will go", "it was torn and green with age", "I tried but he did it".

The governor is always the conjunction; almost any word or phrase can figure as dependent.

2. Adverbial Groups

These consist of an adverb of some kind (i.e. any word which can be a dependent in a group of type 4) together with another adverb or qualifying phrase.

E.G. "almost exactly", "truly in my opinion"¹, "far but not too far out".

The governor is the adverb qualified; the dependent is the word or phrase which qualifies it.

3. Adverbial Clauses

These consist of a verb-form appropriate to this function accompanied by a noun as object, and occasionally a subject as well. The verb-form usually used in English is the participle in "-ing".

E.G. "reaching the shore", "considering the circumstances", "they having at last come".

But other forms may be used, including the infinitive with "to" (e.g. "to tell the truth") and the imperative.

The governor is the participle or other verb-form; the dependents are the noun group(s).

4. Adjunct Groups

These consist of an adjective or verb (without auxiliaries) accompanied by an adverb or group of type 2 or 3.

E.G. "very good", "discredited only in parts", "go after two days".

The governor is the adjective or verb; the adverbial or qualifying elements are the dependents.

5. Participial Clauses

These consist of a participle accompanied by a noun-group; many of them can also function as type 3 groups.

E.G. "eating breakfast"; (This is type 3 in "eating breakfast we waited for the post" but type 5 in "two people eating breakfast"). An example of one which cannot ever be adverbial in function is hard to find, but in practice it is usually safe to ignore the adverbial use except for phrases habitually used as clause-qualifiers (like "considering the circumstances"). The great majority of words will be found to have the same digit in the third and fifth place in the participation class.

The governor is the participle; the noun group(s) are dependent(s).

6. Relative Clauses

These consist of an ordinary verb or predicate together with a relative pronoun (the accusative of the relative pronoun is usually zero).

E.G. "who gave it me", "who I gave it to".

The governor is the predicate; the relative pronoun is dependent.

7. Nominal Groups

These consist of a noun, pronoun or numeral together with any number of adjectives, articles, etc.

E.G. "the three white and slightly battered pigeons that arrived".

The governor is the principal noun (there may be no actual noun owing to ellipsis or quotation: "the big is better than the small", "the now has no duration"). All the other components are dependents.

8. Infinitive Clauses

These consist of an infinitive or gerund of a verb accompanied by one

or more noun groups.

E.G. "(I asked) to take him in my arms", "swimming in hot water (is tiring)".

These groups also generally have the functions of types 3 and 5 as well.

The governor is the infinitive or gerund; the noun group(s) are dependents.

9. Verbal Groups

These consist of a verb with auxiliaries.

E.G. "have come", "has been being", "could change".

The governor is the auxiliary or auxiliary group; the dependent is the main verb or participle.

10. Predicates

These consist of a verb or verb group plus an accusative noun or pronoun. In English, the accusative case is only marked, except in pronouns, by position, and it is usually convenient to ignore the predicate as a separate bracket-group when the object of the verb is an ordinary noun. Thus the only predicates allowed for in our procedure are those with the pronouns "me", "us", "him", "her", "them".

The governor is the verb or verb group; the pronoun is the dependent.

There can be two dependents, as in "give her them".

11. Marked Clauses

These consist of a free clause preceded by a marking conjunction (which commonly has the prefix digit 2 in case a subjunctive verb is used, as happens in written texts occasionally, because this makes it a wait word (see below)). The marking conjunction determines the function of the whole, but in English this function is always adverbial.

E.G. "if you come", "though it was raining".

12. Free Clauses

These consist of a verb group or predicate, one or more noun groups, and possibly qualifying groups types 1-3 as well.

E.G. "they did it", "the book is in the garden", "stop that".

Interrogatives such as "Is that you?", are not given any special treatment, it being supposed that a query mark will always be present which will serve to indicate the interrogative by itself. In fact this does not happen to occur in our dictionary.

The governor is the verb or predicate; all other elements are dependents.

It is clear that in the theory the participation classes of two substituents with the same function will be the same. The 'standard' participation class for a simple substituent may, however, be empirically varied| firstly, if the grouping possibilities of a word are restricted, and secondly, if a word has more than one function. In the first case some of the positive entries in the participation class will be negative; in the second, the participation class will represent the combination (join) of the participation classes for substituents with the different functions here represented by one substituent. It should be noted that compound substituents, on the other hand, have standard entries only.

Habitat and concord

The classification under substituent types represents the main interlingual part of a syntactic dictionary entry. Subsidiary monolingual information can also be given which is extremely useful. It may be possible to specify

unambiguously the position of a substituent in a group; for example, "the" is always the first member of its group, and "ago" is always the last. This information about a substituent, where it can be given, is called its 'habitat'. As this kind of information cannot always be given, however, it is less useful than participation-class information. It is also, of course, different for each language. Monolingual information about what is often called 'concord' may also be useful, though in English, of course, this is not as important as in inflected languages.

Syntactic dictionary entries

The dictionary entries for simple and compound substituents are identical in format; as, however, the following description is illustrated by the English word dictionary being used for tests, it is given in terms of words only.

The dictionary entry for each word consists of three parts. There is first a serial number; this can in principle be applied arbitrarily, and in the present dictionary is simply the number of the word in alphabetical order written in octal notation. The second and third parts require explanation.

The second part of the entry consists of a letter followed by a quartal digit. The letter gives habitat information as shown in the following list:

A	<u>class</u>	<u>covering</u>	<u>example</u>
A	multifinal	postverbs	"out"
E	unmarked		
F	initial	prepositions etc.	"with"

H	final	auxiliary participles	"having"
K	special initial	relative pronouns	"who"
L	special cases (see under)		
M	prefixes		"un~"
O	multinitial	miscellaneous	"all"
R	special initial	articles	"the"
S	medial	conjunctions	"or"
U	special initial	auxiliary verbs	"was"
Z	suffixes		"-ing"

The letter L is used to mark all words which embrace such a set of alternative uses that, we expect, correct analysis will usually be impossible unless they are first sorted out. The list of such words is quite short, and those included are as often as not mere puns which would perhaps cause little trouble in a spoken context. Thus, the fact that "back" can be both a noun and a postverb is sufficiently confusing to warrant its being given the prefix L. In addition, in the present dictionary, all words ending in "ed" or "-ing" are marked L; properly this prefix should be attached to the suffixes themselves; but as we are not here making use of a chunking procedure, the ambiguities of the suffix are carried over to the whole word.

The quartal digit following the prefix letter is used as follows:

- 1 denotes a singular verb or plural noun;
- 2 denotes a constructive verb (i.e. a verb which can take a clause with "that" as direct object); OR,
if the word is not a verb, the digit 2 denotes

- i) accusative case of relative pronouns;
 - ii) Initial conjunctions and other 'wait' words;
- 3 denotes a singular constructive verb;
- 0 denotes any other case.

The 'wait' words mentioned above are words which are liable to form the first members of separated bracket-groups. Thus, the group "either or" is equivalent to a single conjunction, and we therefore count "either" as a wait word and mark it with a 2. No other case of a wait word occurs in this dictionary, but several others exist in English. A special case, which we have not treated in this way, is that of verbs which can take a postverb; in the present procedure postverbs are regarded as qualifying the whole predicate (though this is admittedly not very satisfactory), so that we need not classify all these verbs as wait words. The accusative of a relative pronoun is also a wait word, in that a preposition belonging to it is likely to be found at the end of the clause.

The remainder of the dictionary entry consists of a sequence of the twelve quartal digits constituting the participation class, after the third and eighth of which we put a space. The code for each digit is:

- 1 denotes that the word can be a dependant in the group;
- 2 denotes that it can be a governor in the group;
- 3 denotes that it can be either governor or dependent;
- 0 denotes that it can be neither (i.e. does not occur in such groups in any capacity).

III. 15

Sample dictionary entry

the word "these"

Prefix

Letter	Is the word in any way restricted as to the position it can occupy in its bracket group?		
	NO	The letter is	E
Digit	Is the word a singular verb or plural noun?		
	YES	(The digit is odd)	
	Is it a constructive verb or any kind of wait word?		
	NO	The digit is	1

Participation class

Digit 1	Can the word be dependent or governor of a conjunct group?		
	DEPENDENT ONLY	(e.g. "these and those")	1
Digit 2	Can the word be dependent or governor of an adverbial group?		
	NEITHER		0
Digit 3	Can the word be dependent or governor of an adverbial clause?		
	DEPENDENT ONLY	(e.g. "remembering these")	1
Digit 4	Can the word be dependent or governor of an adjunct group?		
	GOVERNOR ONLY	(e.g. "precisely these")	2
Digit 5	Can the word be dependent or governor of a participial clause?		
	DEPENDENT ONLY	(e.g. "those having these")	1

Digit 6	Can the word be dependant or governor of a relative clause?				
	NEITHER				0
Digit 7	Can the word be dependent or governor of a nominal group?				
	EITHER	(e.g. "these men".			3
		"all these")			
Digit 8	Can the word be dependent or governor of an infinitive clause?				
	DEPENDENT ONLY	(e.g. "to have these")			1
Digit 9	Can the word be dependent or governor of a verbal group?				
	NEITHER				0
Digit 10	Can the word be dependent or governor of a predicate?				
	DEPENDENT ONLY	(e.g. "wanted these")			1
Digit 11	Can the word be dependent or governor of a marked clause?				
	NEITHER				0
Digit 12	Can the word be dependent or governor of a free clause?				
	DEPENDENT ONLY	(e.g. "these will do")			1
<u>Resulting entry; E1 101 21031 0101</u>					

The tests now in progress make use of a small dictionary of five hundred English words. The following is a typical series of entries:

a	001	RO	100	00010	1000
abstract	002	EO	101	21231	3303
abyss	003	EO	101	01021	0101
accomplished	004	LO	100	20210	3202
accordance	005	SO	200	00000	0000
activities	006	E1	101	01021	0101

adopted	007	LO	100	20210	3202	
adored	010	LO	100	20210	3202	
age-old	011	EO	100	20010	1000	
ages	012	E1	101	01221	2303	
agriculture	013	EO	101	01021	0101	
all	014	OO	100	20010	1000	
allows	015	E3	100	00200	2202	
alphabets	016	E1	101	01021	0101	
also	017	EO	130	10000	0100	not allowing for use as a conjunction.
amount	020	EO	101	21231	3303	
an	021	RO	100	00010	1000	
analysis	022	EO	101	01021	0101	
animal	023	EO	101	21021	1101	
another	024	EO	101	21011	1101	

Bracketting

We have so far discussed grouping, or bracketting, in terms of lattice points and lattice algorithms. We must now show how this works out for actual texts. Given that each substituent type defines a kind of group, it is clear that the fact that a set of substituents can be bracketted will be represented in their respective participation classes by a positive entry for the same substituent type. This in itself, however, is not enough: the items to be bracketted must also be contiguous, and must satisfy the governor dependent relation. The latter means that we can only bracket a group of substituents if one of them can be the governor, and the rest dependents, in the kind of group concerned. The governor-dependent relation thus acts as a restriction on bracketting.

Two points should be noted: 1) As many items as possible can be combined at the same time to form a group. The theoretical discussion of grouping was carried on in terms of pairs of substituents, but this is not a necessary feature of the grouping procedure. A group containing more than two substituents obtained by one bracketting can in fact be obtained in several stages by bracketting pairs of substituents at a time. 2) The substituent types are arranged in a priority order from left to right: that is, we look for groups of kind 1 first. The order loosely corresponds to the lattice structure in that "weak" groups are found first, and full clauses last, but it is essentially a practical device for reducing the amount of effort spent in trying to find brackets: as bracketting is carried out on ever larger units, there is clearly some point in looking for the smallest groups of most closely associated substituents first.

The way in which the information contained in participation classes is used for bracketting can be illustrated as follows:

	A	B
x	+	-
y	+	+

This means that x can belong to a group of kind A, but not of kind B, and that y can belong to groups of kind A and kind B. If x and y occur in contiguous positions in a text and can therefore be bracketted, the resulting group must have the function A.

We will now consider a more elaborate case with governor-dependent information given:

	A	B	C
X	G	-	D
Y	D	D	-
Z	-	-	D

When the governor-dependent rule is satisfied only x and y can be bracketted, to give a group with function A, x and z are both dependents in groups of type C and cannot therefore be combined.

Rule-of-thumb-example

In order to keep the example simple the following modifications of the actual procedure have been made:

- i) habitat and concord information is omitted;
- ii) only 6 substituent types are used;
- iii) the bracketting rules are formulated rather crudely.

The sentence to be bracketted is:

A rather lazy cat chases falling leaves and butterflies; of course these can easily get away.

We will assume that the participation class entry for each word has been obtained by dictionary look-up. The sentence with appropriate entries is as follows:

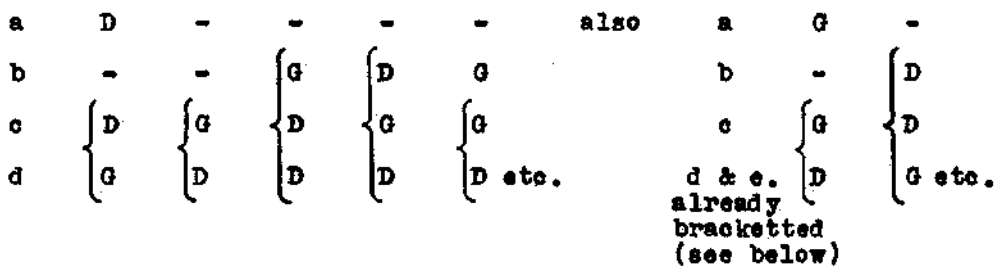
		1	2	3	4	5	6
		C	SA	S	OA	O	Z
a	SA	-	-	D	-	-	-
rather	SB	-	D	-	-	-	-
lazy	SA	D	G	D	-	D	-
cat	S	D	-	G	-	-	D

chases	O	D	-	-	-	G	G
falling	IA	D	D	B	G	D	G
leaves	I	D	-	G	-	G	B
and	C	g	-	-	-	-	-
butterflies;	S	D	-	G	-	-	D
of course	ZA	-	-	-	-	-	D
these	S	D	-	-	-	-	D
can	O	D	-	-	-	G	G
easily	IB	D	D	-	D	-	-
get	OA	D	-	-	G	D	G
away.	OB	D	-	-	D	-	-

Bracketting is carried out according to the following rules;

Rule 1

Starting at the last item before the punctuation stop, whether simple or a compound obtained by previous bracketting (see below), read backwards in each column in turn, looking for the longest continuous sequence immediately preceding the stop in which one item is governor and the rest dependents. As the priority rating of the column is from left to right the first such sequence is taken (even if there is a longer one in a later column). For example, in a particular column the following are all bracket groups:



No brackets starting from c can be obtained in the following:

a	-	D	G
b	D	G	-
c	D	-	D

Rule 2

When Rule 1 suggests a bracket in column 1 if the item marked as governor (i.e. the conjunct substituent) is immediately flanked by two items marked as dependent, treat the three as a group. Thus the first case below will bracket but the second will not:

a	D	D
b	D	D
c	G	D
d	D	G

Rule 3

If under Rule 1 in proceeding backwards from a group already made no brackets can be found, take from the beginning of the existing group the smallest number of items compatible with its remaining a group and try backwards from the last of these before trying again with the reduced following group. This is shown in the following example:

i)	<u>1</u>	<u>2</u>	ii)	<u>1</u>	<u>2</u>	iii)	<u>1</u>	<u>2</u>
a	D	D	a	D	D	a	{D	D
b	D	D	b	D	D	b	{D	D
c	{G	-	c&d	D	D	c	{G	-
d	{D	G	new			d	D	{G
e	{D	D	entry			e	D	{D
			(see					
			below)					

Treatment of bracket groups obtained

The column in which a bracket is made represents the type of the resulting compound substituent. Reference to Table I below gives the appropriate participation class entry, and the group with this new entry is treated as a single item in further bracketting.

Table I: participation class entries for compound substituents of each type:

	1	2	3	4	5	6
	O	SA	S	QA	O	Z
2 : SA	D	-	D	-	D	-
3 : S	D	-	G	-	-	D
4 : QA	D	-	-	-	D	G
5 : O	D	-	-	-	G	G

The participation class entry for a compound containing 1 : C is the meet of the entries for the dependent items.

Bracketting can now be carried out as follows:

Stage 1

By Rule 1 we can bracket in column 4 the last three items:

a	-	-	D	-	-	-
rather	-	D	-	-	-	-
lazy	D	G	D	-	D	-
cat	D	-	G	-	-	D

III. 23

chases	D	-	-	-	G	G
falling	D	D	B	G	D	G
leaves	D	-	G	-	G	B
and	G	-	-	-	-	-
butterflies;	D	-	G	-	-	D
of course	-	-	-	-	-	D
these	D	-	-	-	-	D
can	D	-	-	-	G	G
easily	}	D	D	-	D	-
get		D	-	-	G	D
away.		D	-	-	D	-

By referring to Table I we give this group the participation class entry for a substituent of the type represented by the column in which the bracketting was made, i.e. 4:

D - - - D G

Stage 2

By Rule 1 we can bracket in column 5 the group just made and the preceding item:

a	-	-	D	-	-	-
rather	-	D	-	-	-	-
lazy	D	G	D	-	D	-
cat	D	-	G	-	-	D
chases	D	-	-	-	G	G
falling	D	D	B	G	D	G
leaves	D	-	G	-	G	B

and	G	-	-	-	-	-
butterflies;	D	-	G	-	-	D
of course	-	-	-	-	-	D
these	D	-	-	-	-	D
can	{	D	-	-	-	G G
easily get away.		D	-	-	-	D G

We give this group the participation class entry:

D - - - G G

Stage 3

By Rule 1 we can bracket in column 6 the group just made and all the preceding items back to the stop-point marked by the semi-colon

a	-	-	D	-	-	-
rather	-	D	-	-	-	-
lazy	D	G	D	-	D	-
cat	D	-	G	-	-	D
chases	D	-	-	-	G	G
falling	D	D	B	G	D	G
leaves	D	-	G	-	G	B
and	G	-	-	-	-	-
butterflies;	D	-	G	-	-	D
of course	{	-	-	-	-	D
these		D	-	-	-	D
can easily get away		D	-	-	-	G G

This is correct as the group formed is of the type complete clause.

Stage 4

Re-starting from the semi-colon, by Rule 2 we can bracket the last three items:

a	-	-	D	-	-	-	
rather	-	D	-	-	-	-	
lazy	D	G	D	-	D	-	
cat	D	-	G	-	-	D	
chases	D	-	-	-	G	G	
falling	D	D	B	G	D	G	
leaves	{	D	-	G	-	G	B
and		G	-	-	-	-	-
butterflies		D	-	G	-	-	D

We give this group the participation class entry:

D - G - - D

Stage 5

By Rule 1 we can bracket in column 3 the group just made and the preceding item:

a	-	-	D	-	-	-	
rather	-	D	-	-	-	-	
lazy	D	G	D	-	D	-	
cat	D	-	G	-	-	D	
chases	D	-	-	-	G	G	
falling	{	D	D	B	G	D	G
leaves and butterflies		D	-	G	-	-	D

We give this group the participation class entry:

D - G - - D

Stage 6

By Rule 1 we can bracket in column 6 the group just made and the two preceding items:

a	-	-	D	-	-	-
rather	-	D	-	-	-	-
lazy	D	G	D	-	D	-
cat	}	D	-	G	-	-
chases		D	-	-	-	G
falling leaves and butterflies;		D	-	G	-	-

We give this group the participation class entry:

D - - - - G

Stage 7

By Rule 1 this group will not bracket with preceding items:

a	-	-	D	-	-	-
rather	-	D	-	-	-	-
lazy	D	G	D	-	D	-
cat chases falling leaves and butterflies	D	-	-	-	-	G

By Rule 3 the first item in this group will bracket in column 3 with the preceding item:

III. 27

a	-	-	D	-	-	-
rather	-	D	-	-	-	-
lazy	}	D	G	D	-	D
cat		D	-	G	-	-
chases		D	-	-	-	G
falling leaves and butterflies		D	-	G	-	-

We give this group the participation class entry:

D - G - - D

Stage 8

By Rule 1 this group will not bracket with preceding items:

a	-	-	D	-	-	-
rather	-	D	-	-	-	-
lazy cat	D	-	G	-	-	D
chases falling leaves and butterflies;	D	-	-	-	-	G

By Rule 3 the first item in this group will bracket in column 2 with the preceding item:

a	-	-	D	-	-	-
rather	}	-	D	-	-	-
lazy		D	G	D	-	D
cat		D	-	G	-	-
chases falling leaves and butterflies;		D	-	-	-	-

We give this group the participation class entry:

D - D - D -

Stage 9

By Rule 1 this group will not bracket with the preceding item:

a	-	-	D	-	-	-
rather lazy	D	-	D	-	D	-
cat chases falling leaves and butterflies	D	-	-	-	-	G

By Rule 3 the first item in the following group will bracket in column 3 with this group and the preceding item:

a	{	-	-	D	-	-	-
rather lazy		D	-	D	-	D	-
cat		D	-	G	-	-	D
chases falling leaves and butterflies		D	-	-	-	-	G

We give this group the participation class entry:

D - G - - D

By Rule 1 we can bracket in column 6 the two groups:

a rather lazy cat	{	D	-	G	-	-	D
chases falling leaves and butterflies		D	-	-	-	-	G

This is correct as the group formed is of the type complete clause.

We now have the whole sentence bracketted as follows:

```
((a(rather lazy) cat) (chases (falling (leaves and butterflies;))))  
(of course these (can (easily get away.)))
```

References

1. A.F. Parker-Rhodes and M. Masterman : "A New Model of Syntactic Description", Cambridge Language Research Unit, M.L.. 142.
2. A.P. Parker-Rhodes et al : "The Cambridge Language Research Unit Computer Program for Syntactic Analysis", Cambridge Language Research Unit, M.L. 136.
3. V. Yngve : "A Model and an Hypothesis for Language Structure", Proceedings of the American Philosophical Society, Vol. 104. No. 5. 1960.