Notes on transitivity and theme in English

Part 2

M. A. K. HALLIDAY
Department of General Linguistics, University College, London W.C.1
(Received 22 April 1967)

4. INFORMATION

4.1 *Transitivity, mood and theme.* Part I of this paper (sections 1–3) was an attempt to sketch some of the principal syntactic options, having the clause as point of origin, that are available to the speaker of English for the representation of processes and relations, and of objects, persons &c. as participants in them.[1] The term 'transitivity' was used as a general label for this area of grammatical selection. Part II (sections 4–7) is concerned with another range of grammatical options, also associated with the clause, for which 'theme' is being used as the cover term.

The English clause, it is suggested, can be regarded as the domain of three main areas of syntactic choice: transitivity, mood and theme. Transitivity is the set of options relating to cognitive content, the linguistic representation of extralinguistic experience, whether of the phenomena of the external world or of feelings, thoughts and perceptions. Mood represents the organization of participants in speech situations, providing options in the form of speaker roles: the speaker may inform, question or command; he may confirm, request confirmation, contradict or display any one of a wide range of postures defined by the potentialities of linguistic interaction (Halliday, 1967b). Theme is concerned with the information structure of the clause; with the status of the elements not as participants in extralinguistic processes but as components of a message; with the relation of what is being said to what has gone before in the discourse, and its internal organization into an act of communication (cf. the 'organization of utterance' as a syntactic level in Daneš, 1964). None of these areas of meaning is restricted to the clause; but for each the clause provides a significant range of options, and it is these clause options for which the terms 'transitivity', 'mood' and 'theme' are here being used: given the clause as domain, transitivity is the grammar of experience, mood is the grammar of speech function and theme is the grammar of discourse.

Each of these three sets of options specifies a number of syntactic functions, or roles, combinations of which make up structures of the clause. Some transitivity roles, such as actor, goal and range, were discussed earlier, and these will be returned to and revised in Part III (cf. also Halliday, 1967c). The

roles specified by the theme systems are of a different kind; here structural function is function in communication, and one of the functions is itself labelled 'theme' (section 5), so that the term 'theme' is being used both as a general name and as the name for a particular role in the distribution of information in the clause as message: it is hoped that the context will always make it clear in which sense the term is being used. Other functions within this area are labelled by such terms as 'given' and 'new'; 'topic' and 'comment' are however avoided because they have tended to be used in a way which conflates what are here regarded as distinct functions, with 'topic' meaning both 'given' and 'theme'.

Within the theme system complex, six distinct but related sets of options will be recognized. Three of these concern the status of single elements in the clause structure and these will be treated briefly in section 7; the other three relate to the clause as a whole or, in one case, a distinct unit of comparable extent, and these will be discussed in sections 4–6 under the headings 'information', 'thematization' and 'identification'. In general terms, options in the first set are realized by phonological features of intonation, those of thematization by the sequence of elements in the clause and those of identification by certain specific patterns of clause structure. It is the first of these, the information options, that are not strictly speaking clause systems, since they define their own domain of operation; but since the present paper is concerned primarily with the clause the discussion will be mainly limited to the distribution of information with the clause as starting point.

4.2 *Information unit.* Any text in spoken English is organized into what may be called 'information units'. The distribution of the discourse into information units is obligatory in the sense that the text must consist of a sequence of such units. But it is optional in the sense that the speaker is free to decide where each information unit begins and ends, and how it is organized internally; this is not determined for him by the constituent structure. Rather could it be said that the distribution of information specifies a distinct constituent structure on a different plane; this 'information structure' is then mapped on to the constituent structure as specified in terms of sentences, clauses and so forth, neither determining the other.

Information structure is realized phonologically by 'tonality', the distribution of the text into tone groups: one information unit is realized as one tone group. It is noticeable that in modern English there are two fairly distinct tendencies in punctuation: some writers tend to punctuate according to the information structure, others more according to the sentence structure. The distinction is sometimes referred to as 'phonological' (or 'phonetic') versus 'grammatical' punctuation; but this is perhaps misleading since the two represent rather different aspects of the grammatical structure, the former being 'phonological' in the sense that, since information structure is realized directly

in the phonological organization, it can be interpreted as a marking off of phonological units.

Since the clause is the point of origin for other thematic systems one may set up a 'marked / unmarked' form of option in the mapping of information structure on to sentence structure: in the unmarked option one information unit is one clause. More specifically, it is one non-embedded clause together with all clauses embedded within it.[2] But the information unit may be less than a clause or more than a clause or any combination of these, for example the whole of one clause and part of another; all these are regarded as marked. Put the other way round, a clause may be realized as a single tone group ('unmarked tonality'), or as two or more tone groups, part of a tone group &c. ('marked tonality'). So, for example, the following are all possible as variants of the (written) clause *John saw the play yesterday*, representing different options in respect of information structure:

// John saw the play yesterday //

//John // saw the play yesterday //

// John // saw the play // yesterday //

//John saw the play yesterday but said nothing about it //

// John // saw the play yesterday and is seeing it again today //

The first is unmarked; the clause is one information unit. The remainder are marked, with the clause as two information units, three information units, part of an information unit and one information unit plus part of another respectively; and other variants are also possible. The present discussion will take account only of those options where the clause is organized into a whole number of (one or more) information units; this makes it possible to consider the distribution of information within the general framework of thematic systems, taking the clause as point of origin.

The choice of 'information distribution', with the clause as point of origin, is thus in the first instance a numerical one: how many units of information? The answer will be a whole number from 1 to n, where n is the number of ultimate constituents in the clause. In the case of continuous informal discourse the value of this variable is usually very low; it rarely exceeds the total number of clause constituents – that is, of the elements of structure of the non-embedded clause and of any clauses embedded within it. If one takes the total distribution into account, including information units which extend across clause boundaries, the average number of information units per clause lies between 1 and 2. The following is an example from a recorded conversation:

// I had one of those nice old tropical houses // I was very lucky // it was about thirty years old // on stone pillars // with a long stone staircase up // and folding doors // back on to a verandah // and I came through the door from the kitchen // and a thief carrying my handbag // emerged // through my bedroom door // into the living room // at the same moment //

The distribution into information units represents the speaker's blocking out of the message into quanta of information, or message blocks. Each information unit is realized as one tone group, in the sense that the information structure specifies the boundaries of the tone group to within certain limits, its exact location being determined by considerations of phonological structure. It may be that, above the information unit, it is possible to recognize a higher unit of information structure realized in terms of patterns of tone group sequences specified by tone. This seems more clearly to be the case in loud-reading, where sequences of tone 3 or 4 followed by tone 1 seem to form a clearly marked message unit, often co-extensive with the clause; it is less obviously true in conversation, where the organization of information above the information unit may need to be conceived rather in terms of anaphoric and other cohesive relations between sentences (Hasan, 1967). For the moment we may consider the relation between information units as one of simple linear sequence; this includes the possibility of interpolation (but not embedding), as in

// John saw ⟨// or said he was going to see//⟩ the play yesterday //

where one tone group is enclosed within another.

The linearity of information distribution is in fact related to the internal structure of the information unit, whereby the speaker organizes the components of the message block in such a way as to specify its relation to what has preceded (see 4.3 below). The information unit is what the speaker chooses to encode as a unit of discourse; the decision is a meaningful one, and a text may be structured into such units in any number of ways all other features remaining constant. At the same time the information unit is the point of origin for further options regarding the status of its components: for the selection of points of information focus which indicate what new information is being contributed. The distribution into information units thus determines how many points of information focus are to be accommodated, and specifies the possible limits within which each may be located. It does not however fully specify its location; the assignment of information focus is a distinct option within the information unit. Thus for example (using bold type to indicate information focus)

// **Mary** // always goes to **town** on Saturdays //

contrasts with

// **Mary** // always goes to town on **Saturdays** //

the distribution into information units being unchanged. (It also contrasts with //**Mary** always goes to // **town** on Saturdays // where a different distribution into information units has nevertheless allowed the information focus to be located at the same points. But the interpretation of information focus depends on where it is located relative to the information unit, so that it is the distribution that partially determines the focus and not the other way round. Thus while // on **Saturdays** // Mary always goes to **town** // is normal, // on **Satur-**

days Mary always goes to // **town** // is unlikely, since there is an incongruence between the treatment of *on Saturdays* as theme (see section 5) and its not being assigned the status of a full information unit.)

The distribution of the discourse into information units is related to the options of thematization and identification; this will be discussed in the relevant sections below (5 and 6). It is also related to other, non-thematic options, since it defines the domain for a large number of choices of mood that are realized by tone, as well as being associated with certain features of sentence structure (of which some instances are mentioned in 4.5 below). At the same time it represents a distinct dimension of structural organization, one that is not derivable from other syntactic features.

4.3 *Information focus*. The system of information focus is thus dependent on the information structure; it involves the selection, within each information unit, of a certain element or elements as points of prominence within the message. Each information unit has either one primary point of information focus or one primary followed by one secondary. The choice is again realized in the phonological structure, by the assignment of the tonic (tonic nucleus) in the tone group.

As already noted, the tone group is a phonological unit that functions as realization of information structure. It is not co-extensive with the sentence or the clause or any other unit of sentence structure; but it is co-extensive, within limits determined by the rhythm, with the information unit. The internal structure of the tone group likewise reflects the potentialities of organization within the information unit. The tone group may be considered to consist of one obligatory component, the 'tonic segment', and one optional component, the 'pretonic segment'; as the names imply, the latter, if present, precedes the former. The tonic segment may be either simple or compound; if simple, it has one tonic component, if compound it has two. This determines the range of possible tones; a tone group with simple tonic segment has tone 1 2 3 4 or 5, one with compound tonic segment has tone 13 or 53 (thus the second, or minor, tonic always has tone 3). The initial syllable in the tonic component, the 'tonic syllable', is phonologically prominent, this prominence being primarily a matter of pitch (pitch movement, not pitch level) and secondarily one of duration and intensity. Each segment, and each component within the tonic segment, consists of at least one foot, and therefore contains at least one salient syllable (more exactly, at least one ictus, normally realized as a salient syllable but potentially also as silence).

The system of information focus specifies the structure of the tone group, determining the number and location of the tonic components. Each point of information focus is realized as a tonic component; this begins on an accented (i.e. potentially tonic) syllable, except in certain types of contrastive focus which may determine an unaccented syllable as tonic, and continues until

terminated by a new tone group or a secondary tonic component. Any material preceding the tonic segment in the tone group is pretonic. Thus in

// 4 **John** //13 saw the / **play** / **yes**terday //

there are two information units; the first has a simple tonic segment, focus on *John*, and no pretonic, the second has compound tonic segment, primary focus on *play* and secondary on *yesterday*, and pretonic segment *saw the*. Here the notation is phonological, with // indicating the tone group boundary, / the foot boundary and bold type the tonic syllable; by a simple convention the same symbols can be used, as in 4.2 above, to represent information structure (since the one determines the other), with // as information unit boundary, the foot boundary omitted, and bold type for the word to which information focus is assigned (and not merely its accented syllable).

A text consisting of *n* information units will thus have not less than *n* and not more than 2*n* points of information focus, and for each information unit there are two options: is there just one primary focus or are there two, one primary and one secondary, and where is the focus assigned? The first is realized as either simple or compound tonic segment, the second as the location of tonic prominence. It is not suggested, of course, that such sets of options are separated from one another in real time in the speaker's planning procedures.

Information focus reflects the speaker's decision as to where the main burden of the message lies. It is one of the many diverse phenomena referred to by speakers of English as 'emphasis'; the term is used to cover most of the types of prominence discussed in these sections. Information focus is one kind of emphasis, that whereby the speaker marks out a part (which may be the whole) of a message block as that which he wishes to be interpreted as informative. What is focal is 'new' information; not in the sense that it cannot have been previously mentioned, although it is often the case that it has not been, but in the sense that the speaker presents it as not being recoverable from the preceding discourse. The focal information may be a feature of mood, not of cognitive content, as when the speaker confirms an asserted proposition; but the confirmation is itself still 'new' in the sense intended. If we use the – admittedly rather inappropriate – term 'given' to label what is not 'new', we can say that the system of information focus assigns to the information unit a structure in terms of the two functions 'given' and 'new'.

It was very early observed that in many, perhaps a majority of, instances in English the tonic falls on the last accented syllable in the tone group. This can be interpreted in the light of the phonological structure of the tone group, with obligatory tonic segment optionally preceded by a pretonic segment, to suggest the generalization that the information unit consists of an obligatory new element, realized as tonic, optionally preceded by a given element, realized as pretonic. In the very broadest terms this generalization stands: the information unit may consist only of material under focus, as in // **John** // above, and where

there is also non-focal material this generally occurs as pretonic, preceding the focal. But this general picture of '(given followed by) new' is only partially valid.

Information structure is one aspect of the thematic organization of discourse; and, as brought out in the work of linguists of the Prague tradition (cf. summary in Vachek, 1966: 88 ff.), thematic organization is clearly reflected in various features of modern English syntax, including certain tendencies of word order. The sequence of elements in the clause tends to represent thematic ordering rather than ordering in transitivity of the 'actor – action (– goal)' type, and this is particularly true of the function of clause-initial position which reflects a division of the clause into 'theme' and 'rheme', with theme always preceding rheme (see section 5 below). The functions 'given' and 'new' are however not the same as those of 'theme' and 'rheme'. The two are independently variable (hence the present avoidance of the terms 'topic' and 'comment' referred to earlier). But there is a relationship between them such that in the unmarked case the focus of information will fall on something other than the theme; it will fall at least within the rheme, though not necessarily extending over the whole of it.

While therefore the given – new structure is not itself realized by the sequence of elements, and the focus of information may fall anywhere in the information unit, the partial congruence between this variable and the one which is in fact realized by the sequence of elements, that of theme – rheme, together with the partial congruence between clause and information unit, results in a tendency towards a left to right form of organization in the information unit with given, if present, preceding new. The phonological organization of the tone group into a tonic segment with optional preceding pretonic segment is thus explicable in terms of the tendency for given to precede new in information structure. But whereas in the clause theme always precedes rheme, the theme–rheme structure being in fact realized by the sequence of elements within the clause, the sequence given – new in the information unit is merely the unmarked sequence; the realization is in terms of tonicity and this sequence is far from obligatory: to cite a text example,

//1 **all** the / G C E / papers have to be / marked out of / two / hundred //

The focus of the message, it is suggested, is that which is represented by the speaker as being new, textually (and situationally) non-derivable information. This is why in types of discourse involving much factually new material, as for example the first sentence of a new topic in a broadcast news bulletin, the speaker or loud-reader tends to opt for a large number of short information units, each with its focus of information – and why those news readers who attempt to imitate a more conversational intonation pattern, appropriate to a register with less lexical content per unit grammar and more anaphora, are often quite hard to follow. But the non-predictability of the new does

not necessarily imply factually new information; the newness may lie in the speech function, or it may be a matter of contrast with what has been said before or what might be expected. Nor does the specification of one element as new necessarily mean that all else in the information unit is fully derivable. The interpretation of 'new' and 'given' as functions in information structure, like that of any other structural functions, may vary considerably with variation in the environment. At the same time such variation is predictable from the general interpretation, the significance of which appears in the informational restrictions on anaphoric elements: what is anaphoric by reference is new only if contrastive, while what is anaphoric by substitution cannot be informationally new.

4.4 *Given and new.* In any information unit that is non-initial in a discourse, recoverable information tends to be represented anaphorically, by reference, substitution or ellipsis. Ellipsis involves systemic features having no realization in structure and therefore having no potentiality of association with information focus: what is unsaid cannot be otherwise than taken for granted. By 'reference' is meant here the anaphoric use of what are essentially items of extralinguistic, situational reference (Lyons, 1966: 231), such as pronominals and demonstratives, as in *does John rent this house? – no, he's bought it.* Substitution involves those items which are essentially text-referring like *one* and *do*: *has anyone seen the play? – I think John has done.* From the point of view of information structure WH– items are reference items when interrogative and substitutes when relative.

Anaphoric items are inherently 'given' in the sense that their interpretation depends on identification within the preceding text. Substitutes can in fact never carry information focus; they cannot be structurally new (examples such as *is John going to see the play? – I think he's already done so* are not in fact counter-examples; here it is the tense that carries focus, and *done* must be followed by the non-focal substitute *so*). Reference items however can be structurally new, not only in reference to the situation (i.e. when the deixis is non-anaphoric; for this see 6.3 below) but also when used anaphorically; in the latter case, 'new' is always to be interpreted as 'contrastive', as contrary to some predicted or stated alternative. Thus an item like *yesterday*, interpretable only by reference to 'today', is contrastive if it carries information focus: compare // John saw the play in **June** // with // John saw the **play** yesterday // – where // John saw the play **yesterday** // would imply something like 'not the day before' or 'and therefore doesn't want to see it today' (and see below, on 'unmarked focus'). Similarly // I saw **Mary** // but // I **saw** her // , // I saw **her** // being contrastive; // three months **earlier** // but // three **months** ago//. Compare also the discussion of *he felt himself* in 3.3 above.

Intermediate between reference items and lexical items in general are non-anaphoric closed system items such as verbal auxiliaries and prepositions.

Since prominence in a closed system is inherently contrastive, with such items also information focus implies contrast: // **on** the table // means 'not under, not beside, &c.', // **can** go // means 'not cannot, can but will not, &c.' Thus reference and other closed system items will not carry information focus even when final in the information unit unless they carry contrastive information; this applies even if they are polysyllabic words with an accented syllable, hence // which flag do they **sail** under //, // why don't they **play** together //, // he **hurt** himself //. This is why the rule about the location of the tonic is often formulated as 'the tonic falls on the (accented syllable of the) final lexical item in the tone group', 'lexical item' being understood to exclude closed system items, those which occur as the unique realization of a grammatical feature and thus form one-member classes. Stated in this form, the rule is still not complete, since the tonic may fall anywhere within the tone group; what it specifies is unmarked information focus.

Before explaining what is meant here by 'unmarked', however, it will be useful to consider what is the domain of focus in the information unit, since this may include more than one accented lexical item. Consider an example such as

// ˏ I'm / looking for the / caretaker who / looks after / this / **block** //
The domain of focus here is the whole of the clause constituent *the caretaker who looks after this block*, but only the final accented syllable is within the tonic segment. An even clearer example is

// ˏ I / find it / in /compre/**hens**ible //
The tonic falls, in fact, on the last accented syllable of the item under focus, irrespective of the internal structure of this item; compare // the caretaker who looks after this **block** might know where to find them // (and also *the caretaker who looks after this block's new car*, since possessive 's is likewise assigned by position in the syntagm, not by function in the structure). The domain of focus is thus not the tonic component as such but, in general, the highest rank constituent within which the syllable that is tonic is the last accented syllable. This is often a constituent of the clause; but it may be a constituent of the group, as in // I've seen **better** plays //, where *plays* must lie outside the domain – or even of the word, in which case the tonic syllable may be one which is not accented, as in // the damage was only **external**//.

The information focus assigns the function 'new' to what is within its domain. What lies outside that domain can be said to have the function 'given'; and here the distinction arises between unmarked and marked focus. If we consider the implied questions to which the information unit could stand in answer, then

//**John** painted the shed yesterday //

// John **painted** the shed yesterday //

imply respectively 'who painted the shed yesterday?' (or 'did Mary paint . . . .?',

&c.) and 'what did John do to the shed yesterday?' (or 'did John mend . . .?', &c.); and similarly (since *yesterday* is a reference item)

// John painted the shed **yesterday** //
implies 'when did John paint the shed?' (or 'did John paint the shed this morning?' &c.). But

// John painted the **shed** yesterday //
does not necessarily imply 'what did John paint yesterday?' (or 'did John paint the wall . . .?', &c.); it may simply imply 'what happened?'. Whereas the first three are equivalent to *John* (*did*), *he painted it* and *yesterday*, in the sense that these would be the predicted forms if the question had actually been asked, the fourth is not necessarily equivalent to *the shed*. The fourth example is thus regarded as unmarked in information focus.

A specific question is derivable from any information unit except one with unmarked focus; one with unmarked focus does not imply any preceding information, and this is in fact the form appropriate to the first information unit in a discourse. Where the focus is unmarked, in other words, its domain may be the whole of the information unit. An item with unmarked focus may thus be represented as being ambiguous, as having the structure either given – new or simply new. The two tend in fact to be distinguished in informal speech; either by tone: here tone 4 would imply a specific focus on *the shed* while tone 1 would suggest the other interpretation; or by rhythm, the given-new having no salient syllable before the tonic (and therefore no pretonic): compare (*what do they do?*) // ˏ they / teach / **classics** // with (*what do they teach?*) // ˏ they teach / **classics** //.

A distinction may therefore be made between unmarked focus, realized as the location of the tonic on the final accented lexical item, which assigns the function 'new' to the constituent in question but does not specify the status of the remainder, and marked focus, realized as any other location of the tonic, which assigns the function 'new' to the focal constituent and that of 'given' to the rest of the information unit. Marked focus may be focus on a reference or other closed system item, whether final or not, or on a lexical item that is not final; here what is structurally new is informationally contrastive, either (as already noted) within a closed system or lexically, and what is given tends to be anaphorically recoverable (except in the special case of marked focus on an intensive such as *very*; this should perhaps be regarded as a special case of unmarked focus, since it seems not necessarily to define a given – new structure). In particular, this means that anything that is post-tonic in the tone group will in either case tend to be grammatically or lexically anaphoric; this is why no further intonation contrasts occur after the tonic syllable, so that we can talk of the 'tonic segment' extending to the end of the tone group. Pretonically, on the other hand, there may be both 'new' material within the leftward-extending domain of the information focus and material which,

though structurally 'given', is not recoverable by anaphora; hence the optional pretonic segment with its range of secondary intonation contrasts.

Thus the system of information focus can be thought of as introducing a binary pattern of given – new into the organization of information, although it does not impose a binary structure on every information unit, since the element 'given' is optional. Within the new there may be, as mentioned earlier, a point of secondary information focus, always realized as tone 3. This tone is characteristically associated, even as major tonic, with dependent information, incomplete, contingent or confirmatory; as a minor tonic it is particularly frequent on clause-final adjuncts, as in // he's going back up **north** on **Thursday** //, these being typical of secondary information points. It is also used to give partial prominence to an item that has been mentioned earlier, as in ('have a chocolate?')

//13 ⌄ no I / don't really / **like** / **choc**olates //

Thus the 'new' element may itself be a two-part structure, the second point of focus marking information that is either new but subsidiary or given but to be noted. A particular case of this structure is the thematic system of substitution, discussed in 7.2 below.

We may conclude this sub-section with a text example showing the functioning of the given – new structure in a discourse:

A: //1 ⌄ how / **mad**dening and you //1 had to / walk from / one to the / **other** //

B: //4 not only / **that** but you //1 didn't know /where to start / **look**ing for the / other and a//1 **gain** as I //1 **say** //4 since you / couldn't walk / round very / **eas**ily you //1 just didn't / **try** to / look for the / other //

Here the referents of the given elements *other* and *look for* are close at hand, but the point of reference may be at some distance away; the same text had //4 ⌄ I was a / student / **then** // related to the earlier (answer to *I thought it was half price for students*) // it **is** // but I'm not a **student** unfortunately //, from which it was separated by 83 information units.

4.5 *Some particular cases.* A number of specific contrasts in English grammar are derivable as special cases from the general pattern of information structure. One example is the distinction between defining and non-defining elements attached to nominals: in // his brother the **heart** surgeon //, one information unit, *the heart surgeon* is defining, while in // his **brother** // the **heart** surgeon//, two information units (with tone concord), the two nominals are in apposition. This is the same distinction as that between defining and non-defining relative clauses; here information structure determines sentence structure. But these and similar cases are essentially no different from other instances of choice in information structure, where the distinction between one information unit and two appears merely to reflect the decision as to what information is to be given prominence: for example ('what did you do yesterday evening?') // we took a

boat trip // up the **river** // as far as **Richmond** //, contrasting with // I had dinner with **Evelyn** // where having dinner is a predictable evening occupation. These will not be pursued further here; but we may mention finally one or two points relating more specifically to questions of theme.

The information unit frequently defines the domain of constituents whose status in sentence structure does not fully specify a domain for them. For example, clause-initial adjuncts normally have as their domain the whole of the information unit immediately following: either that in which the adjunct itself occurs or, if the adjunct forms an information unit on its own, the next one. Thus in

//4 ⌄ for / **some** / reason he's //1 gone a/way and he / hasn't / left an ad/**dress** //

he hasn't left an address is part of what is unexplained: it is within the domain of *for some reason*; while in

//4 ⌄ for / **some** / reason he's //1 gone a/**way** and he //1 hasn't / left an ad/**dress** //

only *he's gone away* is within the domain of *for some reason*. The contrast would hold whether or not *for some reason* formed a distinct information unit. The information unit has this function with respect to all adjuncts in this position including those of time, place and so on: compare *yesterday John promised to come but he didn't*, where if there is an information unit boundary after *come* the *but he didn't* is outside the domain of *yesterday* (cf. below, 5.3).

Other, non-initial items such as *only*, *either* and *too* have their domains defined in the same way: compare // Yorkshire gained twelve points and won the **championship** // **too** // where the domain of *too* is the whole of the preceding information unit and the utterance therefore presupposes a preceding one, with // Yorkshire gained twelve **points** // and won the **championship** // **too** // where *too* relates *won the championship* to *gained twelve points* and no previous utterance is presupposed. Here again the role of the information unit in specifying the domain of such items is merely a special case of its function in structuring the message; compare also its role in the specification of layering in co-ordinate structures of three or more elements.

To give an example relating to information focus, the echo question shows the role of information focus in marking a particular option, here an option of mood within the interrogative. In echo questions the WH– element retains its normal function as theme (cf. 5.1), but the focus of information is marked, falling on the WH– element instead of on the final lexical item:

//2 **what** was the / name of the / speaker //

The meaning is 'I've forgotten (didn't catch, don't believe) the answer', and this is achieved by marking the question element as contrastively new, giving it 'thematic prominence' (7.1), with the usual implication that everything else in the information unit is recoverable – that is, in this instance, that the question

has been asked (or implied by being answered) before. Other examples are provided by systems of polarity and contrast in the verbal group, for example // he **took** it // unmarked positive, // he did **take** it // marked positive, // he **did** take it // contrastive positive.

The suggestion is, then, that a piece of discourse consists of a linear sucession of message blocks, the information units, realized by tonality: that is, as a sequence of tone groups the location of whose boundaries is specified to within significant limits by the information structure. Each information unit is the point of origin for the choice of information focus, by which one element is selected as focal, optionally followed by a further, secondary point of focus; this choice is realized by tonicity, the structuring of the tone group into a tonic segment, simple or compound, optionally preceded by a pretonic segment.[3] Information focus assigns the structural function 'new' to a constituent in the information unit, with, optionally, a remainder having the function 'given'. In the unmarked case the new is, or includes, the final lexical item, so that the unmarked sequence, excluding anaphoric elements, is given preceding new; but the focus can appear at any point in the information unit. The constituent specified as new is that which the speaker marks out for interpretation as non-derivable information, either cumulative to or contrastive with what has preceded; the given is offered as recoverable anaphorically or situationally. These are options on the part of the speaker, not determined by the textual or situational environment; what is new is in the last resort what the speaker chooses to present as new, and predictions from the discourse have only a high probability of being fulfilled. Nevertheless the structure of the information unit does contribute in large measure to the organization of discourse, by providing a framework within which these options are exercised.

5. THEMATIZATION

5.1 *Theme and rheme.* It was suggested in the preceding section that the information systems assign to the discourse a structure which is independent of sentence structure and through which the speaker both organizes the act of communication into a chain of message blocks, the 'information units', and specifies within each message block the value of the components in the progression of the discourse. Information is a discourse pattern in the sense that, although the speaker is operating, here as elsewhere in the grammar, with a wide range of options, the factors that he takes into account in exercising these options are those of the textual environment, the preceding discourse; information is thus closely bound up with cohesive patterns such as those of substitution and reference.

Thematization falls within the same general area of options in the grammar of communication; but it is a choice of a different nature. Its point of origin is the clause, not the information unit; and it assigns to the clause a structure in

terms of the functions 'theme' and 'rheme'. It is thus basically a form of the organization of clause constituents, although as will be shown in sections 6 and 7 these can be regrouped by other thematic systems (in the broad sense of 'theme') in various ways which are relevant to the theme – rheme structure. Basically, the theme is what comes first in the clause; and while this means that as pointed out in 4.3, there is in the unmarked case (i.e. if the information structure is unmarked) an association of the theme with the given, the two are independent options (cf. Firbas, 1964). The difference can perhaps be best summarized by the observation that, while 'given' means 'what you were talking about' (or 'what I was talking about before'), 'theme' means 'what I am talking about' (or 'what I am talking about now'); and, as any student of rhetoric knows, the two do not necessarily coincide.

The information systems, in other words, specify a structural unit and structure it in such a way as to relate it to the preceding discourse; whereas thematization takes a unit of sentence structure, the clause, and structures it in a way that is independent of what has gone before. This structuring is into two parts, a theme and a rheme, and is realized simply by the sequence of elements: the theme is assigned initial position in the clause, and all that follows is the rheme. Thus in *John saw the play yesterday*, *yesterday John saw the play* and *the play John saw yesterday* (as a complete clause) the themes are, respectively, *John*, *yesterday* and *the play*.

The theme is what is being talked about, the point of departure for the clause as a message; and the speaker has within certain limits the option of selecting any element in the clause as thematic. As often in this area of the grammar, however, one of the options is unmarked, and the identification of the unmarked option provides a useful insight into the meaning of the theme – rheme structure as a whole.

Which element in a clause would be the unmarked theme depends on the mood of the clause. Let us consider the three mood features declarative, polar ('yes/no') interrogative and non-polar ('wh–') interrogative, as in *John saw the play*, *did John see the play?* and *who saw the play?* (or *what did John see?*). These differ as regards the selection of an element as clause-initial: subject, finite element of predicator, and wh– element whatever its modal function (i.e. whether subject, complement, adjunct or any constituent thereof). In each case, however, the element occuring in first position is the natural theme of the clause.

In a non-polar interrogative, for example, the wh– item is by virtue of its being a wh– item the point of departure for the message; it is precisely what is being talked about. This is why in the unmarked case it occurs initially. The sequence of elements in an English wh– clause is entirely predictable from considerations of theme. Given that *what did John see?* means 'John saw something and I want to know the identity of that something', the theme of

the message is that there is something the speaker does not know and that he wants to know; the rest of the message is explanatory comment about this demand: '(as for) what I want to know (it) is the interpretation of the "something" that John saw'.

The same generalization is valid also for the polar interrogative. Here what is in question is the polarity: the only unknown is whether the text is positive or negative. Since it is the finite element of the verbal group which carries the realization of polarity, this element becomes the theme: in *did John see the play?* the theme of the message is the demand for a resolution of the binary option marked by the *did*. Interaction with information structure introduces many further options relevant to the interpretation of an interrogative but does not alter the thematic pattern. Thus while //2 did John see the **play** // has unmarked information focus, //2 did **John** see the play // has marked focus on *John* and implies something like 'I know Mary did'; but it is still a request for a yes/no answer. Similarly, as noted in 4.5, a non-polar question with marked focus on the wh– item, as in //2 **who** saw the play //, has the special sense of an echo question, but this is merely the result of focus on the theme (cf. //2 **did** John see the play // 'after all? – I know he was going to'), making the request for information itself the point of prominence in the information unit.

In the declarative, the subject is the unmarked theme (cf. Firbas, 1966). The function 'subject' as understood here is specified in the mood systems, not in the transitivity systems; the term therefore corresponds to the 'surface subject' of a transformational grammar, not to the 'deep subject' (which is a transitivity function). The subject is that nominal which, together with the finite verbal element, fulfils a modal role in the realization of speech function; the two together form a constituent specified by the mood systems, one that is obligatory in independent clauses, since these select for mood, but optional in dependent clauses, which do not. Since mood and theme interact here, the subject is also definable as the unmarked theme of a declarative clause.

We can therefore generalize the concept of unmarked theme as the element which the speech function would determine as the point of departure for the clause. In non-polar interrogative this is the wh– item ; in polar interrogative and declarative it is the modal constituent, that containing the subject, and more specifically, within this, the subject in declarative and the finite verbal element in polar interrogative. These represent, in the unmarked case, 'what the clause is about'; and in the declarative this is often an anaphoric or deictic element (having what Firbas (1964: 272) calls 'the lowest degree of communicative dynamism'). But the selection of the theme is a meaningful option within the clause, and the speaker may select, instead of the unmarked theme, a marked theme as in *the play John saw yesterday* ('but I don't think he's seen the film'). We have suggested that in the interrogative the sequence of elements is itself to be explained by reference to the concept of theme; and this is

further borne out by the fact that a clause such as *where the play did John see?* is impossible: since *where* is already the theme the only possible motive for bringing *the play* in front of *did John see* is removed. At the same time the function of the wh– item is as realization of a modal option, so that while the introduction of a marked theme in front of it is predictably rarer than marked theme in the declarative it is by no means impossible: we may have clauses such as *the play where did you see?* or *yesterday what did you do?* Similarly in polar interrogatives; the ordering of the elements in *has John . . .?* is basically a thematic one, so that, again, *has the play John seen?* is impossible, and it is perhaps not unreasonable to suggest that the preference for the 'inverted' interrogative structure in English, by contrast with a number of other languages which possess this resource of inversion but prefer not to use it, may be due to the relative importance assigned to thematic organization by sequence in the syntax of the English clause. At the same time the constituent so formed, of finite verbal element followed by subject, can as a whole be displaced by a marked theme as in *the play has John seen?, yesterday did John see the play?* In other words, while marked theme in the interrogative is infrequent, since interrogatives have a built-in unmarked theme in the form of a question, or rather of a request for information, the speaker may override this by introducing another element as his point of departure.

In the declarative the thematic pressure on the subject is much less strong, and marked themes are frequent in all registers. Marked theme represents a foregrounding of the speaker's point of departure, and its meaning appears from its tendency of association with a particular information structure. In clauses with unmarked theme there is no particular tendency for the theme to appear as a separate information unit; but such a tendency is very noticeable with marked themes, often with tone 3 (non-contrastive) or tone 4 (contrastive): // these **houses** // my grandfather **sold** //, // **that** // I don't **know** //, // **tomorrow** // John's taking me to the **theatre** //. The first information unit consists of the theme and nothing else; and while this is not uncommon with unmarked theme with marked theme it is the predominant pattern, sometimes reinforced by a silent ictus as in

//4 **that** //1 ⌃ I / don't / **know** //

The theme is set apart in a way which the speaker feels may appropriately be glossed, and sometimes does gloss, by some such locution as 'as far as . . . is concerned'. This happens also with the much rarer interrogatives:

//2 this / **sand**wich does //2 nobody / **want** //

where however the theme takes the tone of the rheme. (An apparent exception is the wh– clause with declarative sequence, such as *John wants what?, John went where?* By the definition above these would appear as having marked theme; but they are in fact not characterized by the information structure associated with marked theme, and they are better regarded as having the

thematic structure of declarative clauses. With tone 2 they are echo questions, and take their thematic structure from the clause to which they are responding; with tone 1 they are peremptory questions, perhaps to be regarded as a blend of declarative and interrogative.) As a typical text example of marked theme in the declarative:

//4 ⌄ the / **last** pair of / shoes I / bought there I //1 wore for / ten / **years** //

5.2 *Theme and the passive.* All syntactic options occur in the environment of other options, and the discussion of any one system is likely to require frequent reference to other systems having the same point of origin. In the treatment of marked and unmarked theme reference has been made to both mood and information structure; but the relation of theme to transisitivity has not so far been considered. This concerns particularly the passive, which can be regarded as an option dissociating the roles of actor and theme while leaving the theme unmarked.

It has been said that, as here defined, the subject is that element which is introduced in the realization of certain features in the system network of mood. It is the role which specifies speech function, first as between indicative and imperative and then, in its interaction with the finite verbal element, between declarative and polar interrogative; non-polar interrogative being specified by the wh—role, which may however be mapped on to that of subject. In transitivity, the subject may be actor, goal, beneficiary or range. (It will be suggested in Part 3 of this paper that 'actor' and 'goal' are not wholly appropriate as transitivity roles in English, but this does not invalidate the present discussion.) In information focus, the subject may be (included within) given or new; in the unmarked case it is within the given. In thematization, the subject may be (included within) theme or rheme; what is unmarked here depends on the mood, but if we restrict the discussion to declarative clauses the unmarked option has the subject as theme

Let us consider that each of these sets of options specifies its own constituent structure, each with its own set of structural roles. Then for example

//1 John's / seen the / **play** //

has four simultaneous constituent structures:

(1) transitivity: actor *John*, process *has seen*, goal *the play*

(2) mood: subject *John*, predicate *has seen the play*

(actually, rather, modal constituent *John's*, residual constituent *seen the play*)

(3) information focus: given *John's seen*, new *the play* (alternatively, since the focus is unmarked, new *John's seen the play*)

(4) thematization: theme *John*, rheme *has seen the play*

These roles are mapped on to one another to form complex structural elements: the element of structure is thus a complex of structural roles. Certain options specify the presence of a particular role in the structure: in transitivity, for example, the feature 'effective' specifies the (potential) presence of the roles

of actor and goal. Others have a mapping function, and these include those of 'voice': 'operative' specifies the mapping of actor on to subject and of goal on to complement (that is, on to a nominal in the residual constituent in mood structure; subject and complement are modally defined roles of nominals having the potentiality of participant roles in transitivity). In the example above, which is operative, declarative, unmarked in focus and unmarked in theme, actor = subject = given = theme.

This particular alignment of roles would probably be generally considered to represent the favourite clause type, at least in effective clauses (see however section 10); reasons have been suggested for regarding the pattern which it exemplifies as unmarked in focus and theme, and it is not unreasonable to represent declarative as 'unmarked in mood' and operative, in the environment of effective, as 'unmarked in voice'. In other words if the unmarked term is selected in each set of options the nominal to which the role of actor is assigned will also be the variable that enters into the determination of speech function and the point of departure, but not the point of information, in the message.

Variation from the unmarked results in different role combinations and consitutes the kind of foregrounding that the speaker thinks of as emphasis. Thus in // these houses were built by my **grandfather** //, which is receptive ('marked in voice'), the actor is focal and the goal thematic and the effect is to 'emphasize' the actor as the point of new information and also to 'emphasize' the goal as what the message is about. It seems appropriate to interpret the feature 'receptive' in the grammar as that option which maps goal on to subject because this is the reason for selecting the option: since the subject is the unmarked theme, the receptive allows the goal to be thematic while remaining, qua theme, unmarked – and the actor either to be absent or, if present, to carry the unmarked focus.

If the roles of actor and theme are combined, therefore, they are likely also to be combined with that of subject: *my grandfather built these houses* will be preferred to *by my grandfather these houses were built* because the reason for selecting the receptive does not obtain. But if the roles of actor and theme are separated, the expected pattern is the receptive *these houses were built by my grandfather*, which leaves the theme unmarked and in which the actor is optional. This is not the only possibility; the alternative is operative with marked theme, *these houses my grandfather built*, and this is likely to occur under either or both of two conditions. The first is if the speaker wishes to foreground the theme by marking it, by dissociating it from the subject; we have noted that a marked theme tends to be a separate information unit, and the expected form will thus be // these **houses** // my . . . . The second is if, given that the actor is to be specified, the speaker wishes to assign information focus to the process rather than to the actor yet without introducing the contrastive meaning of marked information focus: // these houses my grandfather

built // rather than with contrastive focus, // these houses were **built** by my grandfather //. The example is now unlikely because of the collocational probability of *built* in the environment of *house*; but if we change it to

// these **houses** // my grandfather **sold** //

the meaning of the choice – that is, of not choosing receptive in spite of having goal as theme – appears clearly: a marked, informationally prominent theme, and non-contrastive focus on the process. Other than under these conditions, if the goal is the theme of the clause the likely transitivity option is the receptive. This allows the actor either to be non-thematic or, in fact, not to be specified at all; and not surprisingly the great majority of receptive clauses in contemporary English texts have no agent (Svartvik, 1966: 141).

The meaning of the passive in English was said many years ago by Mathesius to be related to questions of theme, and it may be explained in such thematic terms: the speaker selects the option 'receptive' in the transitivity system in order to take as unmarked theme a nominal having a role other than that of actor (one of goal, beneficiary or range), the actor either being unspecified or having unmarked focus within the rheme. The option is that of 'receptive', and this may or may not be passive in the sense of having a passive verbal group; indeed the passivity of the verb is entirely unnecessary as a realization of the receptive, which is why the verbal distinction of active / passive now serves a different purpose, that of distinguishing between orientation to the process and orientation to the agency, as in *the houses sold, the houses were sold*; *the door opened, the door was opened*.[4] We may accept Mathesius' view (Vachek, 1966: 91) that the high frequency of the passive in modern English, again by contrast with other languages having like resources, is related to the thematic organization of the clause and its interaction with other dimensions of structure.

Many other factors are involved in determining the use of the passive; but some of these are derivable from this general picture. For example, as suggested by the reference above to collocational probability, the receptive is more likely, other things being equal, where there is a high degree of mutual expectancy between process and goal, since it is in such cases that, if the goal is thematic, the actor will either be omitted or will carry the information focus: both // the bill's been **paid** // and // the bill was paid by **John** // are more likely than // the bill John's **paid** //. At the same time a pronominal actor, being anaphoric, is much less likely to occur as agent than is a non-pronominal one, because the agent is normally new: // the bill was paid by **him** // has marked and strongly contrastive focus.

We could list a paradigm showing the interaction of operative and receptive with the marked and unmarked options in focus and theme, restricting it to the declarative clause as one information unit:

Marked option:	Example:
(none)	my grandfather sold these houses

voice	these houses were sold by my **grandfather**
focus	my **grandfather** sold these houses
,,	my grandfather **sold** these houses
theme	these houses my grandfather **sold**
voice and focus	these **houses** were sold by my grandfather
,, ,,	these houses were **sold** by my grandfather
focus and theme	these houses my **grandfather** sold
,, ,,	these **houses** my grandfather sold
voice and theme	by my grandfather these houses were **sold**
voice, focus and theme	by my **grandfather** these houses were sold
,, ,, ,,	by my grandfather these **houses** were sold

The unlikelihood of the last three is explained by the observation made above, that marking for voice is in modern English a way of avoiding marking for theme, to the extent that the combination of marked voice with marked theme might perhaps be excluded as ungrammatical. Marked theme includes a number of sub-options dependent on the number of elements in the clause (compare the sub-options within marked focus exemplified here); it is only this particular type of marked theme, with theme mapped on to agent, that would be precluded by the choice of the passive. Whether or not voice is treated as fully determined by the theme systems, it seems important for the grammar to show in some way the interdependence between them.

5.3 *Some problems of theme.* Thematization is being considered here in terms of two options: is the theme unmarked (as described in 5.1) or marked, and, if marked, with which element in the clause is it associated? If the structural element is regarded as being a complex of structural roles, the latter question is asking, in effect, with which other structural roles it is associated; and as with unmarked theme so with marked theme the roles specified by speech function will serve as a basis for discrimination. We may designate the elements of the clause here by the familiar labels subject, complement, adjunct, predicator and conjunction; and so far, except in the special case of the agent which can now be ruled out, we have illustrated marked theme only with thematic complement. Other elements may also be selected as thematic.

Thematic predicators are rare (though see section 6 for thematic nominalizations including the predicator); one might however cite the two types exemplified by *standing outside the door he was* ('large as life!') and *and resign he did!* The finite verbal element, being part of the modal constituent, is not part of the predicator and therefore does not here enter into the theme; moreover in the second type it is the finite verbal element that carries the information focus – contrastively, since it is a closed system item (this being an instance of the verbal polarity system referred to in 4.5). There are two clause types in which the predicator regularly occurs in initial position, namely imperative and non-

finite dependent clauses, but in neither of these is it a marked theme. In the imperative the predicator is the unmarked theme, together with *you* or *let's* if present. Non-finite clauses, which also have no modal constituent, select for theme only in a restricted sense; they will be returned to briefly below.

The most frequent type of marked theme is the adjunct, as in *yesterday John saw the play*. Indeed it may seem questionable to group this together with *the play John saw yesterday* as marked theme by contrast with *John saw the play yesterday* as unmarked. That *yesterday* in the first example is thematic is shown by the fact that it cannot be followed by a thematic complement: *yesterday the play John saw* is as unacceptable as *the play yesterday John saw* (unless, in the last example, *yesterday* is qualifying *the play*, in which case *the play yesterday* is a single element and thus thematic as a whole). But although a marked option is usually less frequent than an unmarked one marked is not to be equated with rare; it simply contrasts with unmarked, an unmarked option being one which is in some respect less motivated than others in the same system and is therefore selected unless there is specification to the contrary, corresponding to the final term in an ordered OR relation in a stratificational grammar (Lamb, 1966). In thematization the unmarked theme is wholly derivable from speech function, from the mood system; while any other option requires positive specification by a motivated choice. In this sense *yesterday* does constitute a marked theme.

It is possible however for more than one adjunct to appear in thematic position in the clause. Such instances may be complex adjuncts with internal hypotaxis or embedding, such as *yesterday before dinner* or *in the desk in the corner*, in which case, like co-ordinate structures, they are single elements of the clause. But examples such as *the other day in Sheffield I watched an interesting new process* (or *in Sheffield the other day . . .*) show that the function of theme, restricted elsewhere to single clause elements, can in the case of adjuncts extend over two or more. Like other marked themes such items are often separated by information structure from the rest of the clause, and they may appear either as one or as two information units: // the other day in **Sheffield** // or // the other **day** // in **Sheffield** //.

It was pointed out in 4.5 that information structure specifies the domain of adjuncts appearing initially, and this can now be seen to be related to their function as theme. It seems to be a property of marked thematic elements that, if they occur as a separate information unit, their domain extends over the whole of the next following information unit; hence a pattern such as // these **houses** // my grandfather sold and the rest of his property he left to **me** // is unacceptable, since *these houses* has no function within the second clause. If they are not so separated their domain extends over the whole of the information unit in which they themselves occur but can be terminated by the occurrence of a contrasting item: // these he sold but those he **didn't** //. Thus the

theme plays a part in the bracketing function of information structure; and this may be a property of theme in general, since a theme which is a separate information unit seems, whether marked or not, to be presupposed in any clause included within the next information unit: while // John saw the play and Mary went to the **concert** // is regular, *Mary* displacing *John* as theme in the second clause, // **John** // saw the play and . . . // implies a following clause presupposing *John* as subject, such as //**John** // saw the play and **liked** it //, so that // **John** // saw the play and Mary went to the **concert** // is unacceptable.

In this sense all adjuncts are 'sentence adjuncts', whether of time / place or of cause, manner, &c., since their domain may extend over various levels of sentence structure. It is the thematic status of such adjuncts that makes this possible; and what restrictions there are are restrictions on their occurrence as theme. With time and place adjuncts there is normally a structural relation among sets of them occurring in the same clause, and this limits their thematic possibilities: we can have *tomorrow he's coming at eight* but not *at eight he's coming tomorrow*, since the relation between *tomorrow* and *at eight* is a directional, hypotactic one. But in *he moved stealthily across the lawn like a cat* any of the three adjuncts could occur thematically, there being no necessary structural relation between any pair of them.

A distinction must be made, however, between adjuncts of time, place and cause, manner, &c. on the one hand and items like *perhaps* and *however* on the other, with reference to the relation between theme and initial position in the clause. The concept of theme, like the other options under discussion, is based on the notion of choice: it represents an option on the part of the speaker, and any clause can be regarded as being in contrast with one or more others differing from it just in the selection of the theme. An item occurring obligatorily in initial position will not, in this sense, be thematic; and a corollary to this is that thematic variation in the clause is possible following such items.

This is the case with conjunctions. Co-ordinating conjunctions, both the 'pure' co-ordinators *and* and *or* and the portmanteau items *but yet so* and *then* (those that contain the component *and*, such that *a b but c* is interpreted as 'a and b, but c', &c.), can be followed by the full range of thematic variation, as in *but the play John has seen*; they are in any case best regarded on other grounds as not being constituents of the clause. Subordinating conjunctions, which also have obligatory initial position, permit at least some thematic variation in the clause: *although yesterday he denied it, . . .* (the accented *although*). But this variation is somewhat restricted, so that these items are felt to be clause constituents and there is a slight thematic flavour about their occurrence in initial position. It is as if one aspect of the theme of a dependent clause was the fact and nature of the dependence; and this is seen also in the non-finite dependent clauses, both those introduced by a preposition such as

without knowing the answer and those not, in which the occurrence of a complement or adjunct before the predicator is extremely unlikely. Probably only independent clauses, that is those which select for mood, exhibit the option of theme in its full interpretation; but this may be accounted for if one says that, while the theme system does operate in dependent clauses, the interpretation of theme in this environment requires the recognition of it as secondary to the underlying theme of such a clause, its relation of dependence to another clause. Intermediate in this respect between conjunctions and the adjuncts of time, place, manner &c. are the two classes represented by *however, nevertheless, in that case, therefore* and by *perhaps, probably, frankly, apparently*. These resemble conjunctions in that they cannot be predicated: there is no *it was however that he came, it is perhaps that he will come*. They favour initial position but are not restricted to it; and this suggests that they are thematic when occurring initially. They can, in addition, be followed by elements that are themselves certainly thematic, as in *perhaps tomorrow we can go out for the day*.

These two groups of items fall into the two areas of the grammar that we have called discourse and speech function respectively: those like *however* we may call 'discourse adjuncts', those like *perhaps* 'modal adjuncts'. They have, when occurring initially, the status of a theme; but it is a theme specified in terms of discourse structure or speech function, and thus does not preclude the enunciation of a further theme in the area of cognitive meaning, such as a complement or an adjunct of time or manner. The only restriction is the natural one that a modal adjunct cannot be thematic if the mood is already marked thematically; this is why *perhaps*, &c., cannot occur initially – though they can occur elsewhere – in an interrogative clause. A clause such as *perhaps after dinner we'll go to the theatre* has thus a composite theme, having both a modal part and a part functioning as theme in the general cognitive sense that we have been considering.

With modal themes we have as it were completed the circle, which began with the discussion of the thematic status of WH– items and the finite verbal element, because it is in the modal sense that 'question' is the theme of an interrogative clause. Here the underlying modal-thematic pattern has been built in to the structural realization of interrogatives in the language. At the same time a WH– or finite verb theme differs from a thematic modal adjunct in that it has not merely modal function; it is also, and simultaneously, a 'cognitive' theme in the sense that it is demanding an answer in cognitive terms. Hence a modal adjunct admits a (following) cognitive theme in addition, whereas an interrogative element can only be replaced by a (marked) cognitive theme, which then replaces it as the theme of the clause.

The sort of pattern that emerges might be represented as follows:

Category	1	2	3	4	5	6	7	8	9	10	11	12
Rheme	saw the play yesterday ...	saw the play yesterday	saw the play yesterday	did John see yesterday?	John see the play yesterday?	saw the play yesterday	John saw the play ...	John saw the play	John saw the play	what did John see?	did John see the play?	John saw the play
Theme · Cognitive · Unmarked · Subject	John	John	John			John						
Theme · Cognitive · Unmarked · Finite verb					did							
Theme · Cognitive · Unmarked · WH–				what								
Theme · Cognitive · Marked · Complement or adjunct							yesterday	yesterday	yesterday	yesterday	yesterday	yesterday
Theme · Non-cognitive · Speech function · Modal adjunct			perhaps						perhaps			
Theme · Non-cognitive · Discourse · Discourse adjunct		however						however				
Theme · Non-cognitive · Sentence structure · Conjunction	although						although					
(and)		(and)	(and)	(and)	(and)	(and)		(and)	(and)	(and)	(and)	(and)

The modal and discourse adjuncts can be represented as being, potentially (i.e. when in initial position), themes of a different kind, here labelled 'non-cognitive', and co-occurring with, not excluding, a 'cognitive' theme. Even subordinating conjunctions may be said to have a kind of thematic status as 'structural' themes, though this is to say no more than to suggest a relation of their obligatory initial position to the meaning of clause-initial position in general. The interrogative themes are labelled 'cognitive' for the reasons given: they can be overridden by a marked cognitive theme because they are themselves 'content' themes, in the sense that they express a request for cognitive information, although it is their modal meaning of 'request' that gives them the status of unmarked themes. We may thus retain the concept of a 'marked theme' as defined in 5.1, this being any element other than that derived from the mood of the clause. The framework suggested is very tentative, and the facts are by no means as discontinuous as they may have been made to appear; there is gradience at many points, and types of interaction that would repay further study in terms of serial relationships (Quirk, 1965).

6. IDENTIFICATION

6.1 *Equative and identifying clauses.* It was suggested in the two previous sections that a text in spoken English is structured simultaneously on the two dimensions of given – new and theme – rheme, the former determining its organization into discourse units and the status of each such unit as a component in the discourse, the latter starting from its organization in sentence structure and framing each clause into the form of a message about one of its constituents, with the further possibility of an optional 'key signature' in the form of a theme relating to discourse or speech function. Of the remaining options in the clause which are related to thematization, in that they introduce further ways of selecting or highlighting a theme or of relating it to the rest of the clause, the most general is that here referred to as 'identification'.

This is the option whereby any clause may be organized into a 'cleft sentence' with equative form, and in a number of possible arrangements. The equative clause, referred to in 3.2, has the form 'x equals y', as in *the leader is John*; the 'equals' relation is an asymmetric one meaning 'is to be identified as'. Any clause can be organized in this equative form through the nominalization of one set of its elements, for example *what John saw was the play*. This can be regarded as a particular form of the organization of information in the clause.

Since it is useful to be able to distinguish terminologically between a clause of this type, with nominalization, which contrasts systemically with a non-nominalized clause, and a simple equative clause without nominalization such as *the leader is John*, the former will be referred to here as 'identifying' clauses: the clause *what John saw was the play* is an identifying clause related to the (non-identifying) clause *John saw the play*. An identifying clause has equative

form, with class 2 *be*, but differs from a simple equative clause in that it represents a particular option in the theme systems.

This is a highly favoured clause type in modern English, not least in informal conversation. Where the non-identifying clause specifies a process and its participants, the identifying clause adds the further information that one of the participants is definable by participation in the process: in *what John saw was the play*, 'the play (and nothing else)' is the exclusive goal of John's perception, as far as the communication situation is concerned. It is not surprising that the London brewer's slogan *we want Watney's*, which envisaged the possibility that we might want other things as well, was very early replaced by the identifying form *what we want is Watney's*; nor that the latter has now survived as a prominent feature of London advertizing for close on two decades.

The selection of the feature 'identifying' assigns to the clause a structure in terms of two functions defined by this relation of identification. The two functions are those of an equative clause: a 'thing to be identified' and an 'identifier', that with which it is to be identified. In an equative clause of the WH– type, the 'identifier' is the WH– item, for example *who* in *who is the leader?* In the answer to this question, *the leader is John*, *John* is the identifier. Thus the identifier, in an equative or identifying clause, is that element which corresponds to the WH– item in the WH– question presupposed by that clause. The two elements may occur in either sequence: the answer to *who is the leader?* may be either *John is the leader*, with identifier preceding identified as in the question, or *the leader is John*, with the sequence reversed.[5]

The identifying clause can thus be thought of as an equative in the systemic environment of a non-equative, the functions identified – identifier being those of the equative structure. The system of identification exploits as a thematic option the structural resources of this particular type of transitivity pattern. But there is a restriction on the distribution of functions in an identifying clause. In an equative such as *John is the leader* either *John* or *the leader* may be the element to be identified: the presupposed question may be either 'which is John?' or 'which is the leader?' In an identifying clause, it is always the nominalization which is 'to be identified'.

It is in fact the nominalization which realizes this function; and in the usual pattern the identifier consists of a single element, the remainder of the clause falling within the nominalization. Thus related to *John painted the shed last week* we may have (cited in the sequence identified – identifier):

(PCA)^S the one who painted the shed last week was John
(SPA)^C what John painted last week was the shed
(SPA)^A when John painted the shed was last week
(SCA)^P what John did to the shed last week was (to) paint it

It is possible to include two or more elements in the identifier provided the predicator is among them:

(CA)^SP what happened to the shed last week was that John painted it

(SA)^PC what John did last week was (to) paint the shed

(SC)^PA what John did to the shed was (to) paint it last week

(S)^PCA what John did was to paint the shed last week

(C)^SPA what happened to the shed was that John painted it last week

(A)^SPC what happened last week was that John painted the shed

(Ø) SPCA what happened was that John painted the shed last week

The remaining members of the paradigm can be filled out only facetiously, e.g. *the one who painted what last week was John the shed*. The identifier may be an adjunct of time, place, cause or manner, the nominalization having WH– or noun head (substitute) form: *how he did it was with a knife, the way he did it was with a knife*; nor is it necessarily a clause constituent: *what I need your help with is the wiring, you're the one whose picture I want*. If the predicator is within the identifier, the nominalization must contain a substitute verb, in addition to the finite element if present: *what John can do is (to) paint the shed, what John has done* . . . not *what John can, what John has*.

There is always only one identified and one identifier, so that in any given clause only one nominalization is derivable from this sytem. A clause such as *what I want is what John wants* is ambiguous, being the identifying clause related either to *I want what John wants* or to *John wants what I want*; so in

what do you want?   (what I want) is what **John** wants

what does John want?   what **I** want is (what John wants)

only the bracketed nominalization, which realizes the identified, is derived from the feature 'identifying'. Clauses of this type are particularly frequent where the subject is a nominalization of 'fact', which in the non-identifying clause would occur initially: *what puzzles me is why he left so early*. Since the nominalization is always the identified, in an interrogative identifying clause the nominalization cannot include the WH– item (except in the special case of second order questions such as *// what* **who** *saw was the play //*): related to *what did John see?* we may have *what was what John saw?* but not *the one that saw what was John?*

Thus while any clause containing a nominalization having the equative function of 'identified' is thematically 'identifying', and contrasts in respect of this option with a non-identifying clause, the presence of a nominalization is not by itself an indication of this feature. Not only may the structure not be equative, as in *what John saw surprised him* or *the one who painted the shed will have to do it again*; even if it is equative the nominalization may be the identifier, as in *('which is John?') John is the one who painted the shed*, where the fact that this is not an identifying clause can be seen from the unacceptability of *John painted the shed* as an answer to *which is John?*

As in an equative cláuse, the sequence is free; thus either identified or identifier may be thematic. What is significant is that, whichever of the two occurs in first position, the whole of that element is thematic. In an identifying clause it is the equative structure which specifies the constituents having function in thematization, so that in *the one who painted the shed last week was John* the theme is the whole of *the one who painted the shed last week*. Likewise if the sequence is reversed: in *paint the shed was what John did last week*, *paint the shed* is the theme. This is a further aspect of the difference between the two versions of the slogan quoted above: in *we want Watney's* the theme is merely *we*, whereas in *what we want is Watney's* it is *what we want*.

This thematic foregrounding is reflected, predictably, in the information structure, where it regularly happens that an identifying clause is structured into two information units, the boundary between them falling, as usual in such cases, at the end of the theme: *// the one who painted the* **shed** *last week // was* **John** *//*, or with sequence identifier – identified *//* **John** *// was the one who painted the* **shed** *last week //*. Here both the identified and the identifier carry information focus. Alternatively, however, the identifying clause may be a single information unit, with only one of the two elements focal.

In this case the focus, as in equative clauses generally, normally falls on the identifier irrespective of the sequence: *// the one who painted the shed last week was* **John** *//*, *//* **John** *was the one who painted the shed last week //*; compare, in answer to 'which is the leader?', *// the leader is* **John** *//*, *//* **John** *is the leader //*. We have noted above that the 'new' can be interpreted as replacing the WH– element in a presupposed WH– question, although this may be overridden by contrastive focus as in *('I know John painted the house, but who painted the shed?') // John painted the* **shed** *//* **too** *//*. Since in an equative clause the identifier is that which replaces the WH– element, this is precisely what would be expected to be the new; the equative structure may thus embody the given – new relation in its simplest form, the new information being what serves to identify the given element. But it is not possible simply to interpret identified – identifier as 'given – new in the environment equative', since the two are independently variable; the association of new with identifier may be overridden in the same way, contrastively as 'instead' ('which is the leader?') *// John is the* **leader** *// (but Bill's the one who does the work)* or 'in addition' *(John was the one who mended the shed) // the one who* **painted** *it was John //* **too** *//*. In fact one of the most frequent types of identifying clause, that with a demonstrative as identifier, under certain circumstances (see 6.3) normally has the focus on the identified: *// that's what I* **meant** *//*.

The congruence of identifier with new can be regarded as the unmarked information focus for equative, including identifying, clauses, focus on the identified being contrastive, as shown in the examples above. Here, as elsewhere, marked focus is associated with tone 1 if the sense is 'in addition' and tone 4 if

the sense is 'instead'; and similarly if the clause is two information units the expected pattern has tone 1 or tone 4 on the identified. It is thus necessary to modify the generalization concerning unmarked information focus (4.4), such that in the environment of an equative clause the unmarked focus will fall on (the final lexical accented syllable of) the identifier. It may seem a little farfetched to suggest that // John painted the **shed** // has unmarked focus on *shed* while // **John** was the one who painted the shed // has unmarked focus on *John*; but this does reflect the way in which the speaker will tend to read out these two sentences if they are presented to him in written form without context. The relation of identified – identifier to given – new may be regarded as another instance where in the mapping of one set of roles on to another there is a congruence between them such that one alignment will be selected in the absence of positive specification to the contrary.

There appears however to be another dimension to the structure of identifying clauses, that referred to briefly in 3.2 by the terms 'value' and 'variable'. This is relevant to their interaction with thematization and information, and it may be helpful to attempt to bring it into the picture at this point. The starting point is again the equative clause, and the discussion is extremely tentative; much more study is required of equative clauses as such, not confined to a consideration of their relevance to questions of theme.

6.2 *Identified – identifier and variable – value.* The following is a slightly expanded version, omitting the intensive form, of the table in 3.2 above (Part I, p. 69):

Presupposed question	Coding	Operative		Receptive	
(2) which is John?	decoding	John's the **leader**		the **leader**'s John	
C P S		ID/VR	IR/VL	IR/VL	ID/VR

'there's John; which one does he represent?', i.e. find a value for the variable 'John'

(3) which is John?	encoding	the **leader**'s John		John's the **leader**	
S P C		IR/VR	ID/VL	ID/VL	IR/VR

'there are some people; which one represents John?', i.e. find a variable with the value 'John'

(4) which is the leader?	decoding	the leader's **John**		**John**'s the leader	
C P S		ID/VR	IR/VL	IR/VL	ID/VR

'there is the leader; which one does he represent?', i.e. find a value for the variable 'the leader'

(5) which is the leader?	encoding	**John**'s the leader		the leader's **John**	
S P C		IR/VR	ID/VL	ID/VL	IR/VR

'there are some people; which one represents the leader?', i.e. find a variable with the value 'the leader'

All forms are given with information focus unmarked, on the identifier. In (2)

and (3) *John* is the identified (ID), *the leader* the identifier (IR); in (4) and (5), these roles are reversed. In another respect, however, (2) resembles (4) and (3) resembles (5): the resemblance lies in the kind of identification that is involved. This dimension of structure is referred to here as 'coding', and an equative clause may be either decoding or encoding: either finding a value (VL) for a given variable (VR), or finding a variable with a given value. Thus one element is as it were the realization of the other. The decoding // John is the **leader** // means 'John realizes (can be decoded as) the leader', thus answering a question about John's function (pronominally *which are you?*); to use a linguistic metaphor, it is an observation about the semantics of John. The encoding // John is the **leader** // means 'John is realized by (is encoded as) the leader', answering a question about John's form (pronominally *which is you?*); it is an observation as it were about the phonetics of John. The former identifies John through the role which he occupies, the latter by providing a recognition criterion.

The coding option assigns the roles of value and variable independently of the direction of the identificaton, so that these roles are freely combinable with those of identifier and identified. An equative clause such as *John is the leader* thus has four possible interpretations: as decoding of John, structure ID/VR⁀IR/VL, as encoding of John, ID/VL⁀IR/VR, as decoding of the leader, IR/VL⁀ID/VR, or as encoding of the leader, IR/VR⁀ID/VL. Since the decoding of John is equivalent to the encoding of the leader but with identified – identifier roles reversed, (2) resembles (5) in that in both John 'realizes' the leader, while (3) resembles (4) in that in both John 'is realized by' the leader.

The distinction that is here labelled 'operative / receptive', and thus regarded as equivalent to the active/passive distinction in goal-directed action clauses, relates to the ordering in sequence of variable and value. In the operative, the variable is the subject; and hence also the unmarked theme. There is thus an association of variable – value with theme – rheme similar to that of identified – identifier with given – new: in the unmarked case, the identified is given, the identifier new, and the variable is theme, the value rheme. This is not unrelated to the general meaning of theme; in a sense a theme is a variable to which a value is to be assigned. But, as always, the speaker may exploit the contrastive possibility of not mapping the variable on to the theme; hence to the unmarked, operative corresponds a marked, receptive form.

Since *be* has no passive, the operative / receptive distinction in clauses of this type is realized solely by the sequence, and is thus purely a matter of thematic organization. Or rather (to avoid circularity), it is because of the view of operative / receptive as primarily a thematic choice that we are able to regard value⁀variable as the receptive corresponding to the operative variable⁀value. If this is postulated, then in terms of transitivity roles it is variable and value

(and not identified and identifier) that correspond to actor and goal: 'variable' equals ' "actor" in the environment "equative" '. That the receptive, with value (goal) as theme, is the marked form is reflected by the fact that in decoding, where it has the sequence identifier – identified, the receptive is rare and often seems only marginally grammatical (an example would be (*which are you going to be? the king?*) // no the **king** shouldn't be me // (*it should be someone much taller*)); in the encoding type, on the other hand, where the receptive has the preferred sequence of identified – identifier, both operative and receptive are regular.

The variable would thus, like the actor, be subject if the clause is operative, given the present use of 'subject' as a modal and not a transitivity function. If 'subject' is interpreted as 'deep subject', such that *John* is subject in *the play was seen by John*, then the variable would be subject also in the receptive, which would then have the structure C⌢P⌢S; but this analysis is not appropriate to 'subject' in the modal sense, as is shown by forms such as *is the leader John?* and *the leader is you* (not *the leader are you*). (In WH– equatives the WH– item has its normal initial position, but the subject is still clearly identifiable in forms such as *which am I?*, *who can the leader be?* (decoding operative, encoding receptive), *which is me?*, *who can be the leader?* (encoding operative, decoding receptive).) Hence the rejection of the treatment of receptive // John's the **tall** one // 'John is realized by (recognizable by his being) the tall one' as C⌢P⌢S, referred to in 3.2; *the tall one* is 'actor' (variable) but not subject.

It might be postulated, then, that an equative clause is a two-place predication in which actor and goal are to be interpreted as variable and value; which may be decoding or encoding, and which assigns simultaneously an identified – identifier structure such that if decoding the variable is the identified and if encoding the variable is the identifier. The reason for postulating such a transitivity structure is that the information-type structure in terms of identified and identifier does not by itself account for all the facts, for example the pattern of acceptability in equative clauses of various types. It may be that only the variable can be predicated, so that // it's **John** that's the leader // can mean only 'it's John that has the role of leader', and not 'it's John that you can recognize by his being the leader'; hence the improbability of such an example as *it's the capital of France that's Paris*.

Whether the coding option, if admitted, is present in all equatives is very questionable; if the system is a general one there must be conditions under which it is neutralized, otherwise the problem of explaining why textual ambiguities arise at all – which they not infrequently do – is merely replaced by one of explaining why they do not arise more often. But the discussion is relevant to identifying clauses. These would seem in principle to have the same range of possibilities: a variable – value structure, simultaneous with that of identified – identifier, with the features 'decoding' or 'encoding' specifying the

particular combination. We have seen that the nominalization in an identifying clause is always the identified, so that there would be at most only two possibilities of interpretation for any given identifying clause; thus, related to *John broke the window*, with *John* as identifier:

Coding	Operative	Receptive
decoding	the one who broke the window was/is **John**	**John** was/is the one who broke the window

'you see that person who broke the window; that's John'

encoding	**John** was the one who broke the window	the one who broke the window was **John**

'you know that someone broke the window; that was John'

The same clause but with *the window* as identifier is related to a different set *what John broke was the window*, &c.; and likewise for any distribution of the elements of a clause into identified and identifier.

But it is doubtful whether the decoding form is to be regarded as identifying, in the sense of being agnate to a non-identifying clause: if *the one who broke the window is John* presupposes the question 'who is (what is the name of) that person who broke the window?', the form *John broke the window* is hardly acceptable as an alternative. Similarly with a clearly ambiguous clause such as *what they do not reveal is the source of the difficulty*: only the encoding interpretation is related to *they do not reveal the source of the difficulty*. Decoding clauses, in fact, lie as it were in between encoding equatives and intensive clauses. They are treated here as equative, since the sequence is clearly reversible (the above example is ambiguous in either sequence); but a decoding clause, where the identification is one of definition rather than of specification, is not directly relatable to one in non-equative form.

It seems therefore that identifying clauses should be said to be only encoding, and therefore determined as regards their value – variable structure: the nominalization is always the value. There are pairs of clauses, both identifying in form and the one distinguished from the other by coding, both of which are relatable to a non-identifying clause: *what I was reading were his novels*, decoding 'those things you saw me reading', encoding (which may have *was* instead of *were*) 'those of his things which I was reading', with *I was reading his novels* being interpretable in both senses. Compare *what they sell are/is rejects*: decoding, as in *what they sell are bargains*; encoding, as in *what they sell are/is sports clothes*. Here the non-identifying *they sell rejects* is clearly ambiguous. But in its decoding interpretation it is thematically 'odd' in the way that // **John** broke the window // is odd in answer to 'what is the name of that person who broke the window?'. Likewise in *we heard the overture*: the decoding interpretation 'that's what that was' demands (*what*) *we heard* rather than just *we* as theme – and indeed as an identified; compare *they sell bargains*

which is interpretable only in a decoding sense. As is to be expected, co-ordination between encoding and decoding forms is not possible: (*what*) *they sell* (*are*) *bargains and sports clothes*.

The tentative conclusion would be that the 'identifying' option is a way of representing any clause as an encoding equative, with the nominalization representing the function complex 'value' and 'identified'; a decoding equative, even if it has a nominalization as the identified element, not being regarded as identifying since it has no non-identifying (non-equative) equivalent – or only one that is thematically incongruous. In a decoding equative, in other words, the presupposed question is also in the equative form. The feature 'identifying' may then be said to assign a structure merely in terms of identified and identifier, since in any clause whose equative structure was derived from this option the assignment of the functions value – variable would be predictable therefrom: the identified is always the value and the identifier the variable. Clauses such as *they sell bargains, we heard the overture* (in the sense of 'that piece we heard is the overture'), may be regarded as incongruous in the sense that their underlying transitivity structure is in fact equative (or even, in some cases, intensive). They thus represent some kind of marked option in information structure, the converse, as it were, of identification, whereby the equative (identified – identifier) may take on an actor – process form if the identified embodies such a structure within it. Alternatively, decoding equatives of this type could still be regarded as identifying but with identification as the unmarked option, so that *what they sell are bargains* would be identifying but unmarked, *they sell bargains* non-identifying but marked.

6.3 *Anaphoric identifying clauses.* Probably the items which occur with greatest frequency as the identifier in an identifying clause are the demonstratives, particularly *this* and *that*. The demonstrative tends to occur in thematic position and not to carry information focus, which thus falls on the identified, as in // that's what I **thought** //.

In such clauses the item occurring as the identifier is anaphoric, and therefore cannot be new, unless contrastive. This explains the marked information focus; we have seen that the given – new structure may be incongruent with that of identified – identifier, and where the identifier is anaphoric it clearly is so. But demonstratives are not always anaphoric; they are reference items whose reference may be either situational (cf. 4.4. above) or textual, and if textual may be either backward (anaphoric) or forward (cataphoric). It is their particular reference function that determines the information focus. Since what is referred to anaphorically is 'given', while what is referred to situationally or cataphorically is 'new', demonstratives are normally non-focal when anaphoric and focal otherwise. Thus in // that's what I **thought** // *that* is anaphoric to the preceding text, whereas in // **that's** what I want // the focus shows that the *that* is

referring to something in the situation, unless it is anaphoric but contrastive.

This is a general feature of the demonstratives, not restricted to identifying clauses; compare // 1 this would be the best **plan** //, referring back, with //13 **this** would be the best **plan** // which (again unless anaphoric but contrastive) is either cataphoric – the speaker is going on to expound the plan – or referring to an object, perhaps a set of drawings, that is under consideration. The distinction in identifying clauses is merely an instance of this: // this is what I **meant** // ('what you've just said'), //13 **this** is what I **meant** // ('this object here' or 'what I'm just going to say'). The distinction does not of course apply to *which*, which is always non-focal except in the special types of question referred to earlier.

There is some specialization of reference between *this* and *that* (and other pairs of demonstratives) based on the proximity system of 'near/far', 'near' being speaker-oriented (inclusive of addressee; cf. *come* in *I'm coming to Paris next week* if the addressee will be in Paris at the time), present or future time and cataphoric, while 'far' is not speaker-oriented, past time and anaphoric; hence *there's the problem* ('I've just told you'), *here's the problem* ('I'm just going to tell you'). There is thus a twofold distinction between anaphoric //1 there's where he went **wrong** // and cataphoric //13 **here's** where he went **wrong** //. But in fact, although *that* is rarely cataphoric, *this* is quite unspecific (being found anaphorically also in Shakespeare); and since either may refer situationally it is information focus that marks the demonstrative as anaphoric or otherwise. Thus the frequency of marked information focus in identifying clauses having a demonstrative as identifier is accounted for by the fact that it is this that shows the demonstrative to be anaphoric.

It is natural that anaphora, which combines the features of referential and given, should play a large part in the organization of discourse; anaphoric reference in identifying clauses is just one instance of this, and such clauses are related in the normal way to non-identifying forms:

	Non-anaphoric (*cataphoric or situational*)	Anaphoric
	// I want **that** //	// I **meant** that //
ID^IR	// what I want is **that** //	// what I **meant** was that //
IR^ID	// **that's** what I want //	// that's what I **meant** //

where // what I **meant** was that // is improbable, unless followed by *too*, because *that*, which is both identifier and non-theme, is nevertheless marked as given. But the identifying forms are probably at least as frequent, in dialogue, as their non-identifying equivalents; the combination of deixis with identification, particularly when the deixis is anaphoric, being highly effective as a form of communication. The speaker represents one part of the message as to be identified, and then identifies it with something that is shown deictically to be recoverable from the preceding discourse. Furthermore since the demon-

stratives may have extended text reference the identifier is not restricted to an element functioning in the clause; any stretch of discourse may participate in the identifying relation in this way.

Identification involving time, place, cause and manner has been mentioned above; extended text is particularly prone to be adduced anaphorically as cause or manner, with || that's why I **wanted** it || tending to be preferred to || I **wanted** it for that reason ||, and so on. Here, especially with cause, situational reference is less likely, though cataphora is normal, hence || why I wanted it was **this** || but rarely || why I wanted it was that || (in || (so) **that's** why you wanted it || that is contrastively anaphoric). With place identification both situation and text reference are normal; the demonstrative may be either *here | there* or (except finally) *this | that,* the former being also used for extended text reference. With time, where reference is situational if 'now' and textual if 'then', the normal form, again except in final position, is *this is when, that was when;* here *now* and *then* occur only contrastively, as in *now is when he's supposed to be here.*

The demonstratives *this* and *that* thus have a distinct range of functions when they occur anaphorically as a thematic identifier, in an identifying clause. Their referent may be any element in the preceding text, or any extended text; and the identified may be a nominalization of any kind. It is only in this function that *this* and *that* can refer pronominally to time and place; and also only in this function can they refer pronominally to a human, as in *that's who I'm looking for.* Reciprocally, it is only as the identified in an identifying clause with *this* or *that* as identifier that *who* can occur, except archaically, in head plus relative function, as 'the one who', corresponding to *what* as 'the one which': *that's who/whose it is* but not *John's who/whose it is.* Probably the TH–WH– pattern is felt to be so closely bonded as to be in effect one element, the systemic relation of *that's what I want* to *I want that* giving *that's what* the appearance of a single constituent, with *that's who* as a parallel form. At any rate pronominal *this* and *that* occur with human reference only in such environments where the humanness is simultaneously specified: obligatorily before *who,* optionally before *the one* (but more readily if the 'human' element is explicit: *that's the person that . . . , that's the one who . . . , among all those boys that's the one that . . .*); and similarly with place, &c., reference as in *that's where I live* – but not *I met that* ('John') or *I live in that* ('London').

6.4 *Other aspects of identification.* The head of the nominalization in an identifying clause is either a general noun, always with definite article, *the thing (that . . .*), &c., or the corresponding WH– item *what,* &c. For the adverbial elements there is a straightforward correspondence *the time (when/that . . .)* to *when; the place, the reason, the way* to *where, why, how.* With the nominals, only *what* occurs among the WH– items, apart from *who* in the restricted environment

already referred to, following a demonstrative identifier; *what* corresponds to *the thing.* But the general nouns fall into two sets, with *thing* appearing in both: (1) *thing, person, kind* &c., (2) *thing, one* (only). The former are the generic members of the major noun classes, and this set includes *time, place, reason* and *way;* the latter are 'pro-' forms for the nominal group classes of definite (*one*) and indefinite (*thing*).

Either type may occur in the context of an identifying clause. With the generic nouns the pattern is *the thing we need is a/the new dictionary, the person we need is a/the driver, the kind I like is the Jaguar,* &c. This class is open-ended and it is here that it is difficult to delimit the range of identifying clauses; perhaps any noun that dominates the identifier in the lexical taxonomy should be admitted, for example *the train I usually catch is the 8.30, the car I like is the Jaguar,* since a noun in this relation is cohesive in discourse: compare || John drives a **Jaguar** || he practically **lives** in the car ||, with *car* obligatorily given. With the 'pro-nouns' on the other hand the pattern is *the thing we need is a new dictionary | a driver, the one we need is the new dictionary | the driver.* Thus *thing,* and the corresponding *what,* may be used with non-human nouns whether definite or indefinite and with indefinite but not with definite human nouns (and therefore not with proper names); compare *what she needs most is a daughter* but not *what she needs most is her daughter | Mary.* With definite human nouns, including human collectives, where the appropriate WH– item would be *who,* there is no WH– form of the identifying clause corresponding to the form with ('pro-noun') *the one,* (generic noun) *the person,* &c.; *the one I want is John, the ones who rejected it were the committee.* This follows from the general restriction whereby *who* and *which* do not occur as the head of a nominalization; being themselves definite they do not accept defining modifiers.

There appears to be little difference in meaning, in those instances where both forms are possible, between the two forms of the nominalization. Where the identifier is anaphoric the distinction is that between text reference and object reference, as in *that's what I meant, that's the thing I meant;* but elsewhere the two seem interchangeable: *the thing | what I'm worried about is the interview.* One might expect *thing* to imply some measure of concreteness or discreteness, but there is little sign of this in usage; it is regularly used where the identifier is an abstract or a mass noun. What is common to both forms is the definiteness of the nominalization; this is a feature of all identifying clauses, since it is what relates them to the non-identifying equivalent, and hence the requirement of the definite article if the nominalization has a noun as head. The definite article is obligatory here because it is the only member of the class of specific determiners that is 'non-selective'; whereas the others, the demonstratives and possessives, are 'selective' in that they contain within themselves the specificity required to define the noun head, *the* indicates that such

specificity is to be found elsewhere in the environment. In general the reference may be anaphoric (previous mention), situational (including generalized situational specificity as in *the truth, the moon*) or cataphoric (a defining modifier); but in identifying clauses *the* is always cataphoric, signalling that, in *the one who saw the play was John, who saw the play* must be interpreted as a defining attribute. Thus the identification of *John* with an entity defined solely by participation in the process in question (the noun head, being generic, acting merely as carrier for the defining modifier) is equivalent to its non-identifying correspondent // **John** saw the play // in respect of all features except the selection of this option; whereas the latter is already marked by information focus as being an answer to 'who saw the play?' the identifying form adds the information that this status is shared by no-one else under consideration, and this is achieved by the cataphoric *the*. A clause such as *that big fellow who broke the window is John* is not identifying, since *fellow* carries other modifying elements. (In some contexts *that* occurs non-selectively, either pronominally (especially in the plural), with *that | those* for *the one(s)*, or less commonly as determiner: *that one who broke the window was John* is in fact ambiguous, being an identifying clause only if *that* is not here a demonstrative.)

Deriving from the feature 'identifying' is a polarity option that is independent of the polarity of the clause: positive *the one who saw | didn't see the play was John*, negative *the one who saw | didn't see the play wasn't John*. Thus the identification, while definite, may be positive or negative; the negative denies the identification, so that the positive represents a deliberate option to assert it, and is interpreted as such. In the non-identifying clause the domain of the negative is realized by tone: //1 **John** didn't see the play // relates to *the one who didn't see the play was John*, //4 **John** didn't see the play // to *the one who saw the play wasn't John*. Compare, again, the negative identifying *the one who broke the window wasn't John*, related to //4 **John** didn't break the window //, with the non-identifying *that big fellow who broke the window isn't John*, where the difference in meaning is clear. Likewise in modality there are options in both the matrix and the constituent clause: *the one who can tell us must be John*. In general the distribution corresponds to the two types of modality, the cognitive (personal, participant-oriented) and the modal (impersonal, speaker-oriented): *John could have done it*, cognitive 'would have been able to' *the one who could have done it is John*, modal 'it is possible that' *the one who did it could have been John* (the former can be interpreted in the latter sense but not vice versa). Identification does not however present an independent choice of speech function; the mood is that of the clause as a whole, though realized in the identifying (matrix) clause: *was John the one who broke the window?*, &c.

Thus identification provides the environment for further options in polarity and impersonal modality, as well as the operative / receptive distinction

realized in the sequence of identified – identifier. It imposes a binary information pattern in which the clause takes the form of an element 'to be identified' being equated with an 'identifier', the former being a value for which the latter is the variable. The identified, which usually but by no means always includes all the elements of the clause except one, is set apart structurally by nominalization and thus operates as a single component of the message: if initial it is thematic as a whole and regularly constitutes a separate information unit. Although the form is equative, the option is one of theme (in the wider sense), being part of the organization of information rather than a cognitive choice; and equative clauses can themselves take this option, as in *the one that's the leader is John*.

7. PREDICATION, SUBSTITUTION AND REFERENCE

7.1 *Predication.* In sections 4–6 we have discussed three systems concerned with the organization of discourse: information, which specifies an information unit with given – new structure; thematization, which assigns a theme – rheme structure to the clause; and identification, which optionally assigns an identified – identifier structure to the clause. In this section we shall make brief reference to three systems which assign partial structures within the clause, providing further options within the same general area. All three relate in one way or another to the choice of theme.

Predication may involve any cognitive theme (cf. 5.4 above) and is exemplified by *it was John who broke the window*; it is thus realized as an equative structure, with *it . . . who broke the window* as identified, *John* as identifier, the relator being again the class 2 *be*. The meaning is thus very close to that of an identifying clause with the sequence identifier – identified, *John was the one who broke the window*, both being related to // **John** broke the window // and differing from it in respect of only one feature. Structurally predication maps the function of identifier on to that of theme, giving explicit prominence to the theme by exclusion: 'John and nobody else' is under consideration. There is however a difference between a clause with predicated theme and an identifying clause, in the meaning of the highlighting involved. In identification the prominence is cognitive: 'John and nobody else broke the window'; whereas in predication it is thematic: 'John and nobody else is the topic of the sentence' (hence the alternative form of predication with *there*; see below). In most instances the two will appear identical, but the difference emerges in such examples as *it was in spite of the cold that he went swimming*, which has no identifying equivalent (the absence of a WH- form shows that it does not make sense to highlight *in spite of the cold* in a coding relation, though it makes perfectly good sense to highlight it as a theme), and in the frequency of predicated theme in WH- clauses and equatives, as in *what was it you wanted, it's John that's the leader* (beside identifying *what was what you wanted, John is*

the one that's the leader where again the coding presents difficulties of interpretation).

Predictably, since thematic prominence is a form of 'new' information, the predicated element carries the unmarked information focus. Again, however, as in identifying clauses, the focus does not normally fall on anaphoric demonstratives or on WH– items: // it's that that I can't **understand** //, // what is it you **want** //; on the other hand it is precisely with echo questions, where the WH– item is focal, that WH– interrogative themes tend most regularly to be predicated, as in //2 **who** was it you were looking for //, the echo question itself being a good example of the meaning of this kind of prominence. Otherwise, marked focus occurs in such contexts as (*have you told John that the window got broken?*) // it was John who **broke** it //, where the discourse yields an incongruence between the new and the thematically prominent comparable to that which arises between new and identifier in equative and identifying clauses. There is the usual pattern of unmarked association: a prominent theme will normally be new, but it may in any instance be given, including all instances where it is anaphoric unless also contrastive.

Any cognitive theme, marked or unmarked, may be predicated, the least likely being anaphoric unmarked themes, the most likely being marked themes of any kind. The difference between *his earlier novels I've read* and *it's his earlier novels I've read* is again one of the type of prominence: the former implies the contrast 'but his later ones I know nothing about' and thus is likely to be two information units with the contrastive tone 4 on the first, whereas the latter is not cognitively contrastive and means simply 'these are the ones I'm talking about', being more probably one information unit with tone 1 or 13. The distinction appears clearly in the interrogative, which in predicated theme clauses questions the identity of the theme and not the cognitive content of the message: *is it his earlier novels you've read?* by contrast with *his earlier novels have you read?* There is only one selection for mood; there is no *was it the play did John see?* (except by assimilation from *was it the play John saw?*).

In polarity, however, as in identifying clauses, both the (constituent) clause and the (matrix) predication select independently: *it is / isn't John who has / hasn't seen the play*. With this dependent system of positive or negative predication the full range of verbal polarity systems is introduced into theme predication; and from it derives the possibility of what looks like theme predication without the theme – where in fact the predication is itself the theme, as in *it isn't that I don't want to, it must be that he's out of town*; here the theme is simply 'positive' or 'negative' together with whatever mood, tense and modality may be incorporated in it. Again, the predication selects independently for modality, *it may have been John who was given a prize* being related to a form with thematic modal adjunct *perhaps it was John who* ... (it probably does not, however, select independently for tense: there seems little

if any difference between *it was going to have been John who was given a prize* and *it's John who was going to have been given a prize*, though there is often some tense assimilation in colloquial speech.)

Two further options arise in association with predication: that between *it was John who* ... and *there was John who* ..., and that between *it was John who* ... and *John it was who* .... The first concerns the specificity of the predication, and was referred to above as showing the difference between predication and identification. With *it* the theme is defined (uniquely specified), with *there* it is described (non-uniquely specified): 'John and no others' as opposed to 'John, possibly among others', *it* and *there* being the cataphoric forms corresponding respectively to the definite and indefinite article. The second would appear to add little more than rhythmic variation, as in *the dog it was that died*. The form *John it was who* ... is however used in speech in a way which suggests that it may be explicitly asserting the thematic status of *John*, the structural mechanism of the predication otherwise making the *it* appear thematic. In fact this *it*, which contrasts only with *there*, is non-anaphoric and never can be thematic; but the anaphoric *it* (that which contrasts with the personal pronouns) is thematic when clause initial, so that the use of *John it was who* ... avoids even the temporary ambiguity that an initial *it* would yield. Be that as it may, 'marked thematic' predications such as *the other one it was I really wanted* are regular in conversation; and when declarative they do not combine with a negative predication, which perhaps further indicates that this option relates to the explicit thematic status of the predicated theme.

The possibility of clauses such as *yesterday it was John who was given a prize*, *that one it was John that wanted*, shows however that predication is not restricted to the element designated as theme in section 5. Any element may take on the status of a predicated theme, and such clauses may reasonably be regarded as having two themes one of which is specified as an identifier; if a clause such as *it was only yesterday that it was that one you wanted* is considered to be grammatical then the option must be available recursively, the constituent outside the predication being the point of origin for a further selection. Apart perhaps from conditional attributes (although even *it's alive that I'm scared of them* seems acceptable), the element that cannot normally be predicated is the verb. The Celtic dialects of English, no doubt (as has often been pointed out) because of the regularly initial position of the verb in Celtic languages, regularly accept verbal 'themes' and also accept their 'predication', but these do not match for tone and are hardly interpretable in the same sense; as was pointed out above (5.3), genuine instances of verb themes are highly restricted.

The other restriction relates to equative clauses. It appears to be impossible to predicate the value in an equative relation, so that *it's the leader that's John* can only be interpreted as 'this is how John can be recognized'. This is

NOTES ON TRANSITIVITY AND THEME IN ENGLISH

related to the alignment of functions in identifying clauses, and means that the identified element can never be predicated; but this is predictable from the fact that what is predicated has the role of an identifier within the predication. The identifier itself can be predicated, as in *it was the window that was what John broke*, and there can be predication within a nominalization, always in the 'marked thematic' form determined by the obligatory position of the relative: *what it was that John broke . . . .* But in a form such as *it was what John broke that was the window*, *the window* can only be (most improbably) the value; and the related clause without the predication is therefore not an identifying one.

7.2 *Substitution.* Substitution is also associated with the theme, being an option whereby the speaker can assign to clause-final position an element which would otherwise appear as unmarked theme, as in *they don't seem to match, these colours.* The substitute form has three variants: *he's always late, John*; *he's always late, John is*; and *he's always late is John.* West Riding speakers will recognize this last as a favourite clause type (a recorded example is *he hardly moves his body at all doesn't that conductor*); other dialects tend to favour the other variants and to use the pattern much less frequently, but it is recognizable as a regular form of organization of the clause in connected discourse.

If we compare two clauses in which a place adjunct varies between initial and final position, a typical intonation pattern is:

//4 in **England** //1 they drive on the **left** //
(or //1 in England they drive on the **left** //)
//13 they drive on the **left** in **England** //

with, in the latter case, the adjunct appearing as a secondary point of information focus. A similar pair is formed by //1 these colours don't seem to **match** // and //13 they don't seem to **match** these **colours** //, the difference being merely that *in England* is, as we have regarded it, a marked theme (and hence is more likely to appear as a separate information unit) whereas *these colours* is an unmarked theme. The substitute form almost always occurs with tone 13, with the minor tonic on the 'delayed' element; other tones are possible, but it is not possible for such a clause to occur as tone 1 with the delayed element carrying the only information focus.

The information structure shows the significance of this pattern. The substituted element is as it were a delayed theme; like the clause-final adjunct it is a secondary information point, but whereas this is the normal value for the adjunct, its appearance as theme being marked, for the subject it is the marked value. In other words, with the line between marked and unmarked theme being drawn at this point, *they don't seem to match, these colours* is to *these colours don't seem to match* as *they drive on the left in England* is to *in England they drive on the left* with the rider that in the latter pair it is the first which is thematically unmarked while in the former pair it is the second. It is notice-

able that the elements which normally occur as secondary information points are precisely those which also occur regularly as marked or unmarked themes, including modal adjuncts as in //13 you'll **meet** him **probably** //.

Substitution thus reverses the normal sequence of theme – rheme and introduces a delayed theme after the remainder of the message. The theme in this case is the subject, which is substituted in the modal constituent by a concord pronoun; this refers cataphorically, although since the typical context for a substitute clause is one in which the subject is partially recoverable, from the discourse or the situation, while still requiring to be specified as secondary information, it appears that the function of the pronouns as general reference items is not irrelevant to the interpretation. The meaning is, as it were, 'first I'll say what I have to say and then I'll remind you what I'm talking about'.

There are some instances, however, in which the substitute form is the unmarked option, as in *it does interest me how memory works, it's quite helpful to have taken that course, it's very amusing what you told me.* That this is the same option is again suggested by the intonation pattern. It might seem that the structural shape, or even simply the length, determines the preference here for the substitute form; possibly these factors play some part, but the main factor is not the presence of a nominalization but rather its structural role. If the nominalization is of the class 'thing', and thus functioning as a participant in the transitivity structure of the clause, the substitute form is till the marked one, as in *it excited him, what he saw* beside unmarked *what he saw excited him* (cf. *what he was sitting on collapsed*). Clauses in which the nominalization is derived from the feature 'identifying' are of this type. But if the nominalization is of the class 'fact', having the structural role of information, the substitute form is the unmarked one: *it worries me to see him so overworked.* (Two apparently similar types are left out here as requiring separate treatment: those involving a nominalization of 'condition', such as *it hurts me when you get angry*, which is substituted even when complement (*I like it when the clouds gather*); and those involving 'report', where the constituent clause functions structurally as text, as in *it was rumoured that he might resign.* In the former, substitution is obligatory; the latter are not in fact substitute forms – or indeed nominalizations – at all.) It may be suggested that when the message contains an element whose structural role is that of 'fact' such an element is likely either, if fully recoverable, to be referred to by an anaphoric demonstrative or, if not, to be delayed, as partially recoverable (otherwise it would not have the status of fact), until after the cognitive content of the message. Such substitution is not obligatory; but if the theme is a nominalization of this type non-substitution appears to be a marked option.

Substitution yields many familiar ambiguities, such as *it's the truth that he's confessed*: substitute theme 'that he's confessed is true', predicated theme 'what he's confessed is true', both of which regularly have tone 13 (a predicated

theme cannot itself be substituted, the two options being opposed in meaning); together with other more obvious ones involving *it* in anaphoric use, such as the last example in the unlikely sense of 'this is the (particular) truth which he's confessed'. There may also be ambiguity arising from the two functions of the nominalization referred to above: *it amazed me what he had in that cupboard*, where the one that is in the marked form ('the thing that . . .') would usually be written with a comma after *me*, corresponding to an optional silent ictus (not obligatory; the two would normally be identical in speech). A substitute clause is in fact very rarely more than one information unit, so that substitution is not a device for segmenting information but rather a means of distributing it into a particular thematic pattern.

7.3 *Reference.* This is a form of pronominal anaphora within the clause, even less acceptable than marked substitution in formal uses of language but nevertheless fulfilling a communicative function in informal discourse; a text example is *Britain it's all roads*. It is a favourite in ballad and mock ballad styles. It is restricted to declarative clauses with a nominal theme, marked or unmarked.

In some instances the use of reference is clearly relatable to length; it is a form of reprise for a long theme, especially one that is not subject, showing its function in the structure of the clause: another recorded example is *the sound that came floating out on the air I didn't know I had it in me*. But its use is by no means limited to such instances; in general it serves as a means of isolating the theme from the remainder of the clause, since the theme is then not required as a participant in the clause structure, and thus to emphasize its thematic status. For this reason it tends to be associated with a theme which is a separate information unit, one whose role in the cognitive structure of the clause is subordinated to its primary function as introducing and providing the setting for the message.

Both reference and marked substitution might be left out of consideration as representing the speaker's afterthoughts, though they are afterthoughts of very different, more or less opposing kinds. But in fact both are fully integrated into the information structure, quite unlike the various interpolations and repetitions which involve hesitation and unfinished or interrupted tone groups. They are in no way distinguished by pause or rhythmic irregularity, and it seems that they belong in the area of speech planning in the same way as do the other more familiar options brought together for consideration here.

7.4 *Summary of section* 4–7. The discussion in sections 4–7 embodies the notion that the grammar of the English clause includes a set of options whereby the speaker organizes his act of communication as a component of a discourse. It is

this set of options that is referred to in the title of this paper as 'theme', in the general sense in which it is being contrasted with two other sets of clause options, those of transitivity and mood.

The speaker assigns to the clause a two-part structure of theme-rheme, the theme taking initial position in the sequence. In addition to optional non-cognitive thematic elements, those of structure, discourse or speech function, there will be a cognitive theme which is either unmarked (subject, WH– item or finite verbal element, according to mood) or marked (any other element), the latter being thereby foregrounded as a point of departure for the message. The theme may be an item which is recoverable from the preceding discourse but is not necessarily so; the selection is independent of the context. Optionally the cognitive theme, whether marked or unmarked, may be given further prominence by positive or negative predication; this makes explicit the assertion that the rheme is (or is not) valid for this particular theme and, if the predication is specific, for this theme only, and regularly though not obligatorily accompanies the marking of the theme by information focus as new or contrastive information. Alternatively, the theme may be given prominence of a different kind by being 'picked up' by an anaphoric pronoun later in the clause, this first isolating it as a point of departure and then specifying separately its transitivity role in the clause. Or, if the theme is not itself a pronoun and thus textually or situationally given, it may be shown to be partially recoverable from the preceding discourse by being substituted by a cataphoric pronoun and delayed to the end of the clause as a secondary information point; this is the normal pattern with themes that are nominalizations of fact.

Optionally, the clause may be organized into a two-part 'identifying' structure with equative form, in terms of the paired functions identified – identifier and, perhaps, value – variable. This asserts that something 'to be identified' is interpretable by reference to, and more specifically as a decoding of, an 'identifier'. The identified is represented by a nominalization which regularly, though not obligatorily, includes all elements in the clause except one; and one whole term in the equation takes on the function of the theme, appearing in first position. This structure is frequently associated with an identifier which functions anaphorically, particularly a demonstrative, thus integrating the clause into the discourse through the identification of the defined participant with one that has been mentioned before.

Simultaneously the speaker maps on to the clause, as defined in sentence structure, a structure of a different kind in terms of information units, by which he organizes the discourse into message blocks and specifies the status of the components of the message as new information or otherwise. In the unmarked case (in informal conversation) the information unit will be mapped on to the clause, but the speaker has the option of making it coincide with any constituent specified in the sentence structure. In particular he may isolate the theme as a

separate information unit, especially if it is a marked theme or if the clause is structured in the identifying equative form.

Within each information unit the speaker selects one, or at most two, points of information focus; phonologically the information unit is realized as a tone group and the information focus as the tonic component. This assigns a structure in terms of an obligatory new element, defined as that which is within the domain of the information focus, optionally accompanied by an element with the function 'given'. In the unmarked case the focus of information falls on the final element in the information unit other than any that are inherently anaphoric; any preceding elements, which will include the theme, are then non-specific since the domain of the focus may extend over the whole of the information unit. Alternatively some other element, one that is anaphoric or non-final, may carry the information focus, in which case it is contrastively new and the remainder of the information unit has the status of given; the effect is to give to the message the implication of being a response to a specific question.

For any clause of average length, therefore, there is a very large paradigm of thematic options. The thematic systems by no means exhaust the resources of discourse organization in English; they are merely those options of discourse organization that have their point of origin in the clause, or in a unit which does not correspond to a unit of sentence structure but which is regularly associated with the clause. Much of discourse structure involves patterns of reference, ellipsis and the like which lie outside the more restricted conception of linguistic structure and whose range extends across the boundaries of recognized structural units (cf. Hasan, 1967); and while some of these can be stated as options associated with constituents in sentence structure not all are derived from the clause, for example systems of deixis in the nominal group. The clause is of interest because it provides, perhaps in all languages, a point of intersection of three sets of options, referred to at the beginning of section 4, associated with experiential meaning, speech function and discourse organization. These three are interrelated in complex ways, and each of them may be found to shed light on the other two.

[1] I would again like to thank those acknowledged at the beginning of Part I (cf. *JL* 3 (1967). 37–81); and also, with particular reference to the present sections, K. H. Albrow, Ruqaiya Hasan and J. McH. Sinclair.

[2] I use 'embedded' in preference to my own earlier term 'rankshifted'. The term excludes clauses in hypotactic relation in the clause complex, so that it requires to be interpreted in the sense of rankshifted. Non-defining relative clauses, for example, are not embedded.

[3] With regard to English intonation as a whole, in general the phonological structure (tonality and tonicity) realizes thematic options, while tone realizes modal options. But some options are on the borderline of theme and mood, and some systems assigned to mood in Halliday (1967*b*) should perhaps be regarded as thematic rather than modal.

[4] Orientation to the agency does not mean that the agent (actor) must be specified; the

majority of Svartvik's examples are 'agentless agentives'. The infrequency of the agent reinforces its high information content as a non-thematic (and usually 'new') actor.

[5] I had for a time (e.g. in 3.2 above, and in Halliday (1967*a*)) used the terms 'known – unknown' for identified (known) and identifier (unknown). The present terms, while more cumbersome, are I hope less confusing in their interpretation in this context.

REFERENCES

Daneš, F. (1964). A three-level approach to syntax. *TLP* **1**. 225–240.

Firbas, J. (1964). On defining the theme in Functional Sentence Analysis. *TLP* **1**. 267–280.

Firbas, J. (1966). Non-thematic subjects in contemporary English. *TLP* **2**. 239–256.

Halliday, M. A. K. (1967*a*). *Some aspects of the thematic organization of the English clause*, Santa Monica: The RAND Corporation (Memorandum RM–5224–PR).

Halliday, M. A. K. (1967*b*). *Intonation and grammar in British English*. The Hague: Mouton (Janua Linguarum Series Practica 48).

Halliday, M. A. K. (1967*c*). *Grammar, society and the noun*. London: University College.

Hasan, R. (1967). *Grammatical cohesion in spoken and written English, Part I*. London: Nuffield Programme in Linguistics and English Teaching (Papers 7).

Lamb, S. M. (1966). *Outline of stratificational grammar*. Washington, D.C.: Georgetown U.P.

Lyons, J. (1966). Towards a 'notional' theory of the 'parts of speech'. *JL* **2**. 209–236.

Quirk, R. (1965). Descriptive statement and serial relationship. *Lg* **41**. 205–217.

Svartvik, J. (1966). *On voice in the English verb*. The Hague: Mouton (Janua Linguarum Series Practica 63).

Vachek, J. (1966). *The linguistic school of Prague*. Bloomington, Indiana: Indiana U.P.