From text linguistics to a human linguistics of texts: the scientific paper as a case study

John Hutchins

1. Introduction

The publication of articles in journals lies at the heart of communication in all the sciences. The scientific article is both the principal means for scientists to report their work to others in the same field and the permanent record of their achievements for both contemporaries and for succeeding generations. The focus of this chapter will be the linguistic processes of writing scientific articles, reading and understanding them[1] within the broad framework of 'human linguistics' (or 'hard-science linguistics') outlined by Yngve (1996). Its aim is to indicate how traditional text linguistics may contribute to, or become part of, a human linguistics of texts.

Although in traditional text linguistics it has been common practice to describe texts and their structures independently of authors and readers, as if they have 'meaning' in themselves – i.e. just as Chomsky talks about the 'ideal speaker-hearer' – there are many discourse analysts who recognise the impact of individual readers, their intentions and their behaviours on the resulting text, e.g. the contributors to the collections by Mann and Thompson (1992) and by Coulthard (1994), teachers of English such as Swales (1990), etc. The linguistic features of texts have long been studied by discourse analysts, and their findings are of obvious relevance. While they would interpret linguistic elements (lexical items, anaphors, etc.) as signs of particular textual relationships or functions, in the present context we understand them as manifestations and triggers of particular conditions and procedures, and as networks ('plexes') for the knowledge and awareness of individual writers and readers. While it is easier (because it is more familiar) to interpret texts as entities in themselves and as objects that contain messages, we must attempt to interpret text features as epiphenomena of behavioural conditions and behavioural stimulants. While basing much of the following investigation upon 'traditional' text linguistic analyses, the aim in this paper is to sketch the foundations for a mode of text analysis within human linguistics.

2. Basic structures of the scientific text

The basic foundations for this investigation have been laid in an earlier article (Hutchins, 2004). In outline, the scientific paper is described as a sequence of sections devoted to the description of a 'state of affairs', a 'problem' or anomaly, a 'hypothesis' or potential 'solution', some tests and results, an 'evaluation', and 'implications'. Three basic types for scientific papers can be identified: Problem-Solution, Hypothesis-Testing, and Methodological. In a Methodological paper, the author states the deficiencies of current methods and proposes a 'new' method, which he will then demonstrate by tests as an improvement (or substitution) for other methods. In a Hypothesis-Testing paper, the author states the 'problematic' area and

[1] This paper was originally written in early 2002 as continuation of an earlier paper (Hutchins 2004). In it I revisit ideas first formulated within a traditional 'linguistics of language' framework (Hutchins 1977a, 1977b). Victor Yngve has been of invaluable help in guiding my struggles to reformulate and organize my ideas; and of course, without his pioneering work (Yngve 1996), the idea of doing it at all would not have been possible.

describes the various proposals that have been suggested, and then examine each for their advantages and disadvantages. In a Problem-Solution paper, the author describes the current situation and the problem he is tackling, and then the methods he will use.

Each may be seen as variants of a general type with the following sequence of sections: Situation, Problem, Hypothesis or Hypotheses of Proposed Method, Tests (experiments), Evaluations (of results), Solution (or decision), Implications (or Further Work). In practice, scientific articles are usually organised more simply: an 'Introduction', which covers statements of the overall 'topic', the Situation, the Problem and the proposed solution (Hypothesis); then a section of 'Experiments and Results', covering Tests, Evaluation, and Solution; and a final section of 'Conclusions' for statements of Implications and Further Work.

Text structures and relationships between sentences and paragraphs are not arbitrary or free choices made by authors. In the case of a scientific article, the author-scientist wants to express (convey) a proposed 'solution' to a 'problem', to make public that part of his knowledge (his plex) concerning this 'problem-solution' space which he believes is not known to others. In order to articulate the 'solution' for a problem, the author must first describe what the problem area; and in order to describe the problem, he must be sure that the potential reader knows what the background situation is. He must begin therefore by establishing a 'common' reference point, a situation known to some extent by the reader, i.e. a part of his knowledge space that he can assume is also present (although differing in details) in the 'plex' of his readers. These readers are expected to be other researchers in the same field who may be expected to have the same background knowledge. They may also be expected to be aware of some of its problematic aspects. The author-scientists begins therefore by making reference to this common 'knowledge space'.

We should distinguish between the total 'knowledge plex' of the author, i.e. what he knows about the Situation as a whole and in general, and that part of it which he chooses to write about in a particular article – which I shall call here the 'discourse plex'. The aim in writing about this part is to communicate something which readers (or some readers) do not know, in particular to communicate a solution to something which is recognised as a problem.

For the author the whole of the 'discourse plex' is known. For the reader the 'discourse plex' of an article is unknown. The author's task is to relate his 'discourse plex' to what he presumes the reader knows already, i.e. to a relevant part of the reader's 'knowledge plex' (and also to other parts of the reader's plex, e.g. concerning general language and cultural knowledge, much of which may be subconscious – hence the difficulties of translation.). Since the discourse plex is an extract or subnet of a large multi-dimensional 'knowledge plex' its articulation in a text necessarily involves time-sensitive linearization. The communication or expression of a discourse plex involves the tracing or following of conditions and procedures through a network, and linearization implies back-tracking, the mentioning again of elements already referred to (in whole or in part).

We may envisage the author's discourse plex as comprising (a) the background knowledge of the subject which the author presumes any potential reader will have, (b) the 'topic' to be written about: the problem and its solution, (c) the experiments to be described, (d) the results of experiments and how they relate to the proposed 'solution', (e) the author's view of the impact his new knowledge and results may have on other scientists-readers. Excluded from the discourse-plex will be any parts of his subject-knowledge not considered to be relevant or necessary for

the articulation of the specific 'topic', and any parts of his experimental activity or investigations that did not lead to any resolution of the 'problem' being addressed – i.e. any false paths and failures. It is common practice for scientists-authors to present investigations in an idealised fashion, as a 'logical' progression of hypothesis, experiment and results, whereas in actuality most investigations are strewn with mistakes, wrong hunches, false starts, interruptions, etc. (as demonstrated by Knorr-Cetina (1981) – these will invariably be absent from the 'discourse plex'.

The author is presented with a succession of tasks, each with its own set of subtasks. First, he outlines what is accepted knowledge in the particular area he is going to discuss, in vocabulary familiar to expected and potential readers. In describing the state of affairs (Situation), he assumes that potential readers share this knowledge (at least in part). The Situation represents a starting point for both author and reader, and may well be explicitly summarised as the 'topic' (of the 'discourse plex'). The statement of a Situation requires a justification (nothing is said without a purpose), and this is the mention of an anomaly or problem in what is commonly known. The Problem may not necessarily be entirely familiar to the reader (hence the author may provide citations), but its broad outlines will be known to those readers who continue. Any potential reader not knowing the background or unaware (or uninterested) in the problem will probably stop reading at this point. Having established (or stated) a problem, there will be a normal expectation for readers that the author will offer a Solution. The author may well do this by putting forward next one of more suggestions (Hypotheses). In order to investigate them, he will next propose some Tests. The reports of these tests constitute the main scientific (descriptive) content of the article, i.e. the actual physical experiments. Description of tests leads to an expectation that the author will make judgements of their contribution to assessing the validity of the hypothesis, i.e. the tests are followed by Evaluations. Finally, any positive Evaluations leads to an assertion that the problem has been solved, i.e. the expectations from an initial Problem statement are fulfilled.

The reader undertakes a comparable succession of tasks. One is to identify what the situation is to which the author refers, with the subtasks of understanding the 'topic statement', understanding the description of the Situation, and relating to his own 'knowledge plex'. Then his task is to understand the nature of the Problem (as stated by the author) and to decide whether the question is of interest to him. In doing these tasks, he acquires certain expectations – primarily that the author will offer a Solution to the problem and that the answer will be relevant to his own research. The satisfaction of this expectation is fulfilled by the task of locating the Solution in the text and by the task of 'absorbing' this new information into his own plex.

In broad terms, we may also distinguish between 'prospective organization' and 'retrospective interpretation'. The author takes account of the expectations of potential (imagined) readers and takes into account their reactions at each stage of the text – anticipating questions and problems, and providing anaphoric lexical connections, etc. The reader engages in 'retrospective interpretation' discovering the connections made by the author, reacting and questioning, and finding whether he agrees or not with the author's conclusions, etc.

3. Microstructure, macrostructure, superstructure

In most theories of text linguistics there are commonly agreed to be three basic aspects of text structure: microstructure, macrostructure, and superstructure (or

schemata) – using van Dijk's terminology (Van Dijk 1977, 1980). Microstructure refers to relationships within sentences and between adjacent sentences or clauses; macrostructure to relationships between blocks of sentences (e.g. sections and paragraphs) and large text segments; and superstructure to the functions of text segments within the text as a whole and its overall organization as a coherent text.[2]

The 'discourse plex' of an author is not to be equated directly with what text linguists call macrostructure or superstructure. The 'discourse plex' is the region of the author's total plex which is isolated or extracted by the author for communication to potential readers. The macrostructure of a text is conceived as a representation of the linguistic 'content' (or meaning) of the 'final' finished text, and the superstructure is the overall organization of the finished text. The 'discourse plex' is the focal area of the author's 'knowledge plex' before he starts to compose – although it may (perhaps often does) change in range and extent in the course of writing. In addition, while macrostructure is often conceived as an 'abstract' or generalization of the text and therefore, to some extent, a linguistic object, the 'discourse plex' is the non-linguistic set of conditions, procedures, etc. relevant to the text (as conceived by its author). Nevertheless, since we do make preliminary notes and sketches for articles and we know (or presume) that we often *think in language*, it may be hypothesised that there could be some similarity between the 'semi-linguistic' (rough sketch or internalised) articulation of part or all of a 'discourse plex' and what text linguists might interpret as the underlying macrostructure of its corresponding text.

From the viewpoint of human linguistics the notion of microstructure and macrostructure may assist the analysis of distinctions between the linguistic behaviours associated with the expression of those parts of the plex brought into the 'domain of control' and the 'long-term' objectives of an author. The former are the active parts of the 'discourse plex' manifested in 'microstructure', the latter are the latent parts of the 'discourse plex' to be activated at a later time. Although the macrostructure of the text linguist is a retrospective 'construction' of the author's 'meaning', it may be presumed that it does reflect in some way the 'content' of a 'discourse plex'.

The superstructure relates most closely to the author's (internal) organization of his communication (e.g. in terms of Situation, Problem, Hypothesis, Evaluation, Solution), to his plans and goals, his intentions and how he may achieve them. The problem-solution framework described above is found in many linguistic investigations of the structure of "research articles." For example, Swales (1990: 127ff.) established a similar four-part structure: Introduction, Methods, Results, Discussion. Of these four parts, the most complex is the 'Introduction'. Swales identifies three basic components with sub-components (some optional) occurring generally in the following sequences (Swales 1990: 137ff.):

(1) Establishing a territory
- Claiming centrality
- Making topic generalizations
- Reviewing items of previous research

Establishing a niche
- Counter-claiming, or Indicating a gap, or Question-raising, or Continuing a tradition

Occupying the niche
- Outlining purposes, or Announcing present research

[2] Distinctions between micro-, macro-, and super-structure are not sharp or even easily detected in practice, but they are useful methodologically.

Announcing principal findings
Indicating structure of article

As in the Situation-Problem-Tests-Evaluation-Solution frame, Swales relates the structure of scientific papers to the activities of the scientist-authors in organising and presenting the results of their investigations. Thus, the first section, 'Establishing a territory' or 'Situation', locates the work as central to the research field, states the overall 'topic' and reviews previous work. The second, 'Establishing a niche' or 'Problem', indicates gaps and raises questions. The third, 'Occupying the niche' or 'Hypothesis', announces what the article is going to cover and what has been found.

4. Lexical indicators of text structure

Indicators of text structure have dual functions. On the one hand, they are means for authors to organise their 'discourse plex' as a linear text (i.e. convert a network of relations and conditions into a sequence of utterances). And on the other hand, they are pointers and triggers enabling readers to complete their tasks of relating what the author has written to their own understanding and of fulfilling expectations raised by the text itself.

For each of the sections (paragraphs) of a scientific text there are overt linguistic signals for the functions and relationships of segments within the total text. As indicated in Hutchins (2004) these signals should be regarded, within human linguistics as triggers or cues used by authors for arousing in readers conditions or procedures similar to those present in the 'plex' of the author (in this case, in the scientist).

In Swales' analysis of the research article, we see lexical items indicating (or marking, or characterising) section components – of the outline structure in (1) above. For example, 'centrality claims' may be expressed (Swales 1990: 144) as:

(2) Recently there has been a spate of interest in how to...
The possibility of... has generated interest in...
The explication of the relationship between... is a classic problem of...
Knowledge of... has a great importance for...

Statements 'making topic generalizations' include statements about knowledge or practice (3a) and statements about phenomena (Swales 1990: 146):

(3) (a) There is now much evidence to support the hypothesis that...
The... properties of ... are still not completely understood.
A standard procedure for assessing has been...
(b) An elaborate system of... is found in the...
There are many situations where...

Statements 'indicating gaps' or 'question-raising' are (1990: 154):

(4) However, the previously mentioned methods suffer from some limitations
The first group... cannot treat... and is limited to...
The second group... is time consuming and therefore expensive...
Both... suffer from the dependency on...

Finally, there are examples of statements for 'occupying a niche' (Swales 1990: 160ff.):

(5) This paper reports on the results obtained...
The aim of the present paper is to give...
In this paper we give preliminary results...

Some of the most familiar of these lexical indicators are those of 'topic statements', as Meyer (1992) has illustrated more fully. Typically, we find in scientific articles, statements such as: *In this paper we report…, We report here…, The purpose of this study was…, In our present work, we study…* Meyer argues that the common use of the word *report* is a signal from the author that he believes that what he is writing is

not speculation, but an objective account of the facts – he is merely a 'reporter', contrary to the evidence of how scientists actually work from hypotheses to demonstrations of putative validity.

5. Lexical indicators of inter- and intra-sentence relationships

Lexical signals are found marking relationships between paragraphs and between sentences in texts. As Winter (1977) has demonstrated, these lexical signals are taken from a relatively restricted vocabulary set. They are used by authors as relational cues (triggers) not only between paragraphs (or larger sections and text segments) but also between sentences and between clauses within sentences: temporal (*before, after, then, afterwards, while, until, since, at the same time, subsequently, …*), spatial (*behind, in front of, beside, next to, to the rear, on the other side, …*), teleological (*because, therefore*), circumstance (*since, in these circumstances*), result (*so, thus*), reason (*for, because, for this reason*), concession (*although, even though, nevertheless, yet,…*), instrument (*by…*[do]*ing, with, thereby…*), condition (*if…then*), correlation (*as…so*), coordination (and), alternation (*or, either, alternatively*), antithesis (*but, on the contrary, by contrast*), unexpected consequence (*but, however*), and so forth.

One of the most influential text theories has been the Rhetorical Structure Theory (RST) of Mann and Thompson. Of particular interest is that they show (Mann and Thompson 1988) not only that intersentence relations such as 'circumstance', 'elaboration', 'concession', etc. are reflected between larger units of texts, but that they reflect authors' intentions (or goals) with respect to the way authors aim to influence and change the attitudes, opinion and actions of potential readers. Relations are defined as operating between a 'nucleus' (N) and a 'satellite' (S), where either N or S can be any length of text (or utterance) from a single clause to a sequence of sentences or even paragraphs. For N and S there are defined specific constraints. For example, the constraints on the relation 'background' are defined as: For many definitions there is a clear reference to changes of state or of knowledge which the author intends to achieve in his readers, with equally clear relevance to the goals of human linguistics in this area of investigation. For example, 'elaboration' is defined as: "S presents additional detail about the situation or some element of subject matter which is presented in N or inferentially accessible in N …", and its 'effects' (i.e. changes of state in readers R) is defined as: "R recognizes the situation presented in S as providing additional detail for N. R identifies the element of subject matter for which detail is provided." In the case of "solutionhood" the 'effects' are defined as: "R recognizes the situation presented in N as a solution to the problem presented in S."; in the case of "motivation": "R's desire to perform action presented in N is increased." Since Mann and Thompson derive these relations and their definitions (in terms of constraints and effects) from the analysis of linguistically articulated relations in a wide range of different kinds of texts, there would seem to be relevant foundations for the testing the findings of RST and other theories of text linguistics within a human linguistics of texts.

6. Lexical cues and text organization

Hoey (1983, 1994) demonstrates that texts (sentence sequences) exhibit a number of recurring patterns of lexical cues, which may be categorized as sequences such as: Condition-Consequence, Instrument-Achievement, Cause-Consequence, Situation-Evaluation, Evaluation-Basis, Problem-Solution. We may illustrate with one of his examples (adapted) from an advertisement (Hoey 1983: 84)

(6) (a) Over 20 million people in Britain wear dentures. (b) Often the major cause of losing teeth is poor oral hygiene, leading to gum disease. (c) Regular toothbrushing helps, but it is only one part of the answer. (d) A toothbrush simply cannot clean the spaces in between the teeth. (e) If these spaces aren't cleaned, plaque builds up in them. (f) Plaque is a sticky film that clings to teeth… (g) To avoid this condition, use Inter-Dens Sticks regularly. (h) They massage the gums whilst cleaning the interdental spaces, removing the plaque… (i) Use Inter-Dens and help keep your natural teeth for life!

The Situation is described in (6a), and given in (6b) as the Consequence (*losing teeth*, *disease*) of a Cause (*poor hygiene*). To avoid the Problem(s) (*dentures, disease*), sentence (6c) offers one Solution (*toothbrushing*), but it is rejected (*only part of the answer*); the Basis for this negative Evaluation appearing in (6d) and (6e) (*cannot clean*, *aren't cleaned*). The latter is another Cause-Consequence relation (*if…then*) introducing a further Basis for rejection (*plaque*), which is then given an Elaboration (6f). In (6g) the writer refers back to the Problem in (6b) (*to avoid this*), and proposes another Solution (*use*…), which in (6h) is given a positive Evaluation (*cleaning the interdental spaces*, echoing *cannot clean the spaces in between teeth* in (6d)). Finally, the Solution is repeated as an imperative (6i).

The occurrence of one member of a clausal relation sets up the expectation of the other member: a Problem anticipates Solution, Evaluation requires a Basis, etc. Equivalently, we may say that particular words anticipate elaborations and resolutions: *difference, comparison, conflict, difficult, unfortunate(ly), danger, risk, drawback, dilemma, tricky, incompatible.* For example (Winter 1977: 33):

(7) Let's compare our newly-weds as possible life partners. There is likely to be conflict between them. We knew him a long time ago before he got married and so we can say that he and his new bride differ radically in their approaches towards life. Whereas he is frugal and deeply religious, she is spendthrift and frankly worldly in all things.

The word *compare* sets up the expectation that the following text will express some differences between the individuals being talked about (theme = *newly-weds*). The second sentence states that the differences will be a Problem (*conflict*). The third sentence emphasises the serious nature of the differences (*differ radically*), and gives the author's Basis for this Evaluation (*we knew him a long time ago*). The fourth sentence expresses the differences using the clausal relation *whereas.* As yet unfulfilled is the author's Solution to the Problem – which may come in subsequent sentences.

The parallelism between the labelling of these sets of lexical cues and the labelling of text sections (segments) – Problem, Solution, Evaluation, etc. – is not, of course, a coincidence. It reflects the fact that authors (writers of all backgrounds) have a common approach – perhaps learned within a particular culture or discipline, rather than common to all communicating individuals – to the presentation of their 'message'. The first need is the present the context, then the problem to be tackled, some possible solutions, and finally the answer. This pattern is seen again and again at the micro-level of clauses, sentences and paragraphs and at the macro-level of sections and texts.

The problem for a text linguistics – whether within a traditional 'linguistics of language' or in 'human linguistics' – is that many of these sentence and text structural relationships are implicit. Authors leave them unstated, in the expectation that readers will correctly infer the intended relationships from knowledge of the context and background – a background which is either assumed to be shared by author and reader or which the author has invoked (or established from previous text) in the reader.

For example, the sentence (8a):

(8a) It started to rain

may indicate a beneficial or a threatening Situation – in itself (8a) is neutral, but in the context of (8b) it is presumed to be beneficial, and in the context of (8c) it is a threat (i.e. a Problem).

(8b) It started to rain. Our crops would be saved.

(8c) It started to rain. We took our umbrellas.

As readers (hearers), we understand such expressions by reference to our common-sense knowledge of reality, namely that rain is wet, that rain can help crops to grow, that getting wet may be undesirable, that crops are grown as food, that umbrellas can protect us from rain, etc. The author of (8b) indicates that rain is a benefit by the lexical signal saved. The author of (8c) indicates that rain was a 'threat' (or problem) by the lexical signal umbrella. He could have be more explicit:

(8d) It started to rain. To avoid getting wet, we took our umbrellas

but this would usually be considered unnecessary, since we know (in our culture) that umbrellas protect us against getting wet. For individuals living in other cultures, where umbrellas protect against the sun (cf. 8e), the explicitness of (8d) may be necessary:

(8e) The sun was getting hotter. We took our umbrellas.

In addition, of course, as in so many aspects of language, expressions of relationships can be ambiguous. For example, *but*:

(9a) He is not dead, but [on the contrary] he is alive.

(9b) My horse is black, but [by contrast] yours is white.

(9c) They set out for Paris, but [however] they did not arrive

In practice (i.e. in the context of real communication between individuals), there is no ambiguity; the shared (or established) background or knowledge of author (speaker) and reader (hearer) means that the exact nature of the relationship is grasped without difficulty and usually immediately. How this is achieved is one of the major tasks for 'human linguistics', and indeed for any theory of language.

7. Thematic progressions

A consequence of the linear nature of spoken and written language is that authors and speakers have to refer more than once to the same elements in their 'discourse plex'. As each subtask (and sub-subtask, etc.) is undertaken by the author he brings into the current 'domain of control' the relevant elements or parts of the 'discourse plex' and any other parts of the 'background context' assumed to be known already. These conditions and procedures are triggered by explicit descriptions if 'new' and presumed to be unknown, unpredictable or unexpected, or by anaphora if 'old', already mentioned or presumed to be familiar or predictable.

In traditional text linguistics, anaphors are interpreted as references back to something already expressed. In a human linguistics of texts they are to be seen as 'shortened' references to already known parts of the discourse plex. The first mention of an element is given by a 'full' reference (or rather one which is sufficient for identification by the reader or listener); subsequent references can be less specific, conveying only distinctive parts of the element. Anaphora refer not only to previously explicitly mentioned 'parts' of the author's discourse-plex, but also to implicitly associated parts (i.e. other conditions or procedures triggered by the focus on a particular 'domain of control')[3]. Thus, e.g., mention of book may trigger 'awareness' of pages, style and author, each of which may then be referred to as if already explicitly mentioned: I *bought a book without noticing that the pages were blank. The style was attractive but the author unknown*.

[3] For a full account of anaphora and reference in human linguistics see Burazer (2004).

It is often observed that readers of text remember not the wording of texts (even partially) but their overall contents. It is particular apparent for spoken discourse, but it is true also for written texts. What is remembered, in text linguistic terms, is not the actual language but the message. What is remembered, therefore, in human linguistics terms is the associations that a text has built in the plex of an individual (the reader) and that can be reactivated by triggers. This is a process that takes place throughout the reading of a text. From many years of studying the mechanisms of texts, Sinclair (1994), has argued that back references (anaphoric and lexical) do not point, as traditional text linguistics assumes, to words, nouns, clauses, sections, etc. of the previous text but to the "state of discourse". Every sentence has to be interpreted on its own: "The previous text is part of the immediately previous experience of the reader or listener, and is no different from any other, non-linguistic, experience." (Sinclair 1994: 16-17). Anaphors are thus interpreted like proper nouns as references to previous experiences – in this case as experiences gathered from the reading of previous parts of the text.

Both in microstructure and in macrostructure there are progressions from what is assumed by the author to be known or familiar to the reader to what the author assumes to be 'new' information. At the microstructural level, this is the well-known theme-rheme structure where themes are signalled overtly by anaphors, definite descriptions, deictic pronouns and expressions, etc., and where rhematic elements may be signalled by indefinite descriptions. In general (but not invariably) definites, anaphors and deictics refer to what is 'old', presumed to be known or familiar, or ascertainable from previous parts of text, whereas indefinites are used for 'new' (rhematic) information. However, definite descriptions, anaphors, pronouns, etc. do also occur in rhemes. In such cases what is 'new' is the assertion of which they form part.

In traditional linguistics, it is said that words, expressions, sentences, etc. refer to external reality, phenomena, etc. It is also said that anaphoric expressions refer to previous text elements. But it should be clear that it is not words, etc. that refer to 'reality' but the author/writer who *uses* words and sentences (or rather produces sounds and marks of certain forms) in order to refer, and that we should therefore say that the writer uses anaphoric expressions in order to indicate that he is referring to something that he has previously referred to. It is not a 'proposition' (some kind of logico-semantic representation) but an 'entity' (Gundel 1985), an identifiable part of observable reality or of mental 'life'. From the viewpoint of successful communication, anaphoric expressions work well if readers are able to identify the 'same' part of reality (entities) that the author intended to refer to. If, however, the reader does not have the interpretative ability or requisite background knowledge to make the 'correct' identification, then the communication fails.

Typically, thematic elements (themes) precede rhematic elements (rhemes), since it is more natural for writers (and speakers) to start from what is known before going on to what may be 'new' to their readers (or listeners) rather than vice versa.[4] It is also more natural for thematic elements to refer back to some elements of the immediately preceding sentence (or more precisely, for writers to use them to create links in order to refer to concepts, events, etc. already referred to in previous

[4] Halliday and Hasan (1976) define 'theme' solely in these terms, i.e. as the starting point of 'base' of a sentence. Here I follow Firbas (1966) who defines theme as the element with the lowest degree of 'communicative dynamism' and normally consisting of 'given' elements – a definition which corresponds to a 'human linguistics' conception of theme as reference to an activated area of the plex in the 'domain of control' which has already been expressed or alluded to.

sentences.) In this way the writer can convey the message by a 'natural' progressive accumulation of 'new' information. In crude terms there are basically two ways a theme may be related to a preceding sentence: either as a repeated reference to elements of the foregoing rheme, or as a repeated reference to some or all of the preceding theme. There are thus two types of 'thematic progression' (Daneš 1974): linear progression (10) and parallel progression (11).

(10)
$$\begin{array}{l} T_1 \longrightarrow R_1 \\ \quad\downarrow \\ \quad T_2\ (=R_1) \longrightarrow R_2 \\ \qquad\qquad\downarrow \\ \qquad\qquad T_3\ (=R_2) \longrightarrow R_3 \end{array}$$

(11)
$$\begin{array}{l} T_1 \longrightarrow R_1 \\ \downarrow \\ T_1 \longrightarrow R_2 \\ \downarrow \\ T_1 \longrightarrow R_3 \end{array}$$

A simple example of a linear progression is:

(12) The boy (T_1) read a book (R_1); it (T_2) was about elephants (R_2); these animals (T_3) live in Africa and India (R_3)

and of parallel progression:

(13) The boy (T_1) came home from school (R_1); first he (T_1) had something to eat (R_2); and afterwards he (T_1) watched television (R_3)

In most cases, thematic progressions are mixtures of these types, e.g. a common example is that of a 'split rheme' (13), which may be illustrated by a paragraph such as:

(14) All substances are divided into two classes: elementary substances and compounds. An elementary substance is a substance which consists of atoms of only one kind… A compound is a substance which consists of atoms of two or more different kinds…

(15)
$$\begin{array}{l} T_1 \longrightarrow R_1\ \ (=R_{1a} \qquad \& \qquad R_{1b}) \\ \_\_\_\_\_\downarrow \qquad\qquad\qquad\qquad \downarrow \\ \downarrow \qquad\qquad\qquad\qquad\qquad \downarrow \\ T_2\ (=R_{1a}) \longrightarrow R_2 \qquad\qquad \downarrow \\ \downarrow \qquad\qquad\qquad\qquad\qquad\qquad \downarrow \\ T_2 \longrightarrow R_3 \qquad\qquad\qquad \downarrow \\ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\downarrow \\ \downarrow \\ T_3\ (=R_{1b}) \longrightarrow R_4 \\ \downarrow \\ T_3 \longrightarrow R_5 \end{array}$$

Both parallel progressions and 'split rheme' progressions can be marked by lexical cues. Thus, parallel progressions can be marked explicitly by: *first, afterwards, furthermore, then, next, finally*, etc.; and 'split rheme' by lexical markers such as: *on the one hand, on the other hand*. But often there are no lexical markers (as in (14).)

In each of these thematic progressions we see that the first sentence provides the starting point or foundation for the following sentences. In this sense, it functions as the theme for the paragraph as a whole. The initial sentence of a paragraph is often referred to as its 'topic sentence', to which following sentences are related either coordinatively or subordinatively. From the viewpoint of the author it is the 'base', the point of departure for the paragraph – or, in human linguistics terms, it is the initiation of a 'domain of control' for the paragraph.

From the viewpoint of the reader, both the 'theme' of a sentence and the 'base' of a paragraph initiate one of more expectations. The reader must expect that if an author mentions again something that is already 'known' he must be intending to same something 'new' or unpredictable – unless he is summarizing what has been already communicated, in which case, he will commonly signal this fact explicitly (using words such as *in brief, to summarise*). In other words, a theme sets up the expectation of a rheme, and a 'base' sets up the expectation of an elaboration, explanation, solution, etc. according to the nature of what elements of the plex ('discourse plex') are highlighted.

This brings us to the macrostructural organisation of text. Here too we find features of thematic progression. Just as sentences have themes and just as paragraphs have 'base sentences', so complete texts may also have 'base paragraphs' or 'base sections', i.e. those parts which establish the background (or foundation) knowledge which the author will elaborate upon in his text. Invariably (inevitably) the 'base' of a text is the initial section (often labelled 'Introduction' by authors) of one or more paragraphs where the writer invokes those entities, objects, concepts, events, etc. that are to represent the context or background (the 'domain of control' in Yngve's (1996) terms) which he presumes the reader (or potential reader) knows already. The subsequent sections (groups of one or more paragraphs) are linked to this opening 'base section' either coordinatively or subordinatively, in the same way as sentences are in an individual paragraph). We may thus extend the notion of thematic progression to relations among paragraphs. Just as the thematic elements of a sentence may be related to either the theme or the rheme of a preceding sentence, so may the 'base sentence' of a paragraph be related to the 'base sentence(s)' of a preceding paragraph or to the rhematic ('new') parts of a previous paragraph. The result is a linear progression or a parallel progression of paragraphs, or indeed a version of 'split rheme' progression.

The progression of text sections such as 'Situation', 'Problem', 'Hypothesis', 'Testing', 'Evaluation', 'Implications' are instances of thematic progressions at a superstructural (super-paragraph) level in so far as elements of later sections refer back to elements of earlier sections – or rather in so far as the author refers again (anaphorically) to entities, concepts, or events that he has already referred to in previous sections. The labelling of sections may be done explicitly by authors, but more usually these are labels used by analysts based on the occurrence of lexical 'signals' indicating text functions and relations (Hoey 1983, 1991).

8. Expectations

Binding together linguistic behaviour at the microstructural, macrostructural and superstructural levels are the broad concepts of 'expectations' and 'anticipations'. An author attributes to his imagined (expected) readers certain experiences, knowledge, opinions and beliefs. He anticipates certain reactions of readers and acts accordingly. Whenever a writer thinks that his text may raise a doubt or leave a question unanswered he tries to provide the information or gives an explanation or elaboration that he thinks the reader is expecting (e.g. as in Winter's example (7) above.) Failure to fulfil an expectation (e.g. withholding an explanation for some action, not spelling out all implications) runs the risk of possible misinterpretation. It may be done safely only if the author has good reason to suppose that the reader shares the author's viewpoint in crucial respects – which the scientific author may well be able to do, since he writes generally for a highly knowledgeable readership.

Anticipatory indicators in sentences can be interpreted as pointers to 'associations' which the reader can expect to be provided by following parts of the text. Thus, the word *problem* (or *difficulty* or *complication*, etc.) in a sentence indicates that the author believes that the associations (plex) developed so far in the text contains 'anomalies' that he intends to resolve in subsequent sections of the text: he creates an expectation in the reader that a 'solution' will be offered.

At the superstructural level, there are sets and sequences of expectations both for the general context of scientific investigation (theorems, experiments, results, reports); there are sets and sequences of expectations for written scientific communication (writing, revising, journal submission, refereeing, revising, making references; and (as outlined above) there are the sets of expectations in the structures of articles: problems, solutions, tests, evaluations, etc.

We may note the similarity to the research on text analysis in the context of 'artificial intelligence' (Schank 1975). Sequences of expectations and actions referred to as 'scripts' were used to aid the understanding of linguistic behaviour in particular situations, e.g. the restaurant 'script': the expected sequence of entering, sitting, ordering, being served, eating, asking for the check, paying, leaving, each stage having characteristic constraints, conditions and procedures, and characteristic linguistic behaviours.[5]

9. Determining topics

The crucial contact-point for the reader of a scientific article is the introductory section (or paragraph) – the 'base section' – where the author reviews the research (the current state of knowledge), points out lacuna, inconsistencies and anomalies, and then states as clearly as possible what he has discovered or concluded from the work being reported. In this section (typically labelled 'Introduction'), it is quite common for the author to state explicitly what he considers to be its 'topic' as a whole, what he contends it to be 'about'.

It is equally true that the reader will not assume that all thematic elements comprise the 'topic'. The reader will make his own judgement of what the text is about. He may agree with the author's 'topic statement' or he may not. It is more likely that he will identify various elements of the text as 'topics'. They are likely to include subjects of particular interest to the reader at the time when he is reading it, e.g. in the case of a scientific article, it might be a particular method of chemical analysis – for such a reader the article may be mainly 'about' this method. It follows also that 'topics' can change over time; what may interest the reader on one occasion may be of no interest to him on a later occasion. And vice versa, what may have seemed irrelevant (or not understood, and therefore not given topic status) when first read may later become of greater interest on later reading. The individual reader's state of knowledge (his plex of conditional properties) change over time – from experience, from learning, and from reading.

It may be noted that the 'topic' of a text is a statement (single sentence) or summary (paragraph) which is some kind of generalisation of the content, i.e. of the 'discourse plex'. It is the product (expression) of the conditional properties and procedures derived from the 'associative network' (plex) present in the author or reader (described above). In the case of the author, this 'discourse plex' has its origin in the actual experiment being described and how the author (as scientist) has reflected on it, and expressed it. In the case of the reader, the network of properties

[5] A human linguistics analysis of the related 'diner' scenario is given by Yngve (1996: 185ff.)

and procedures is the result of his interpretation and understanding of the text. The generalisation of 'topics' from such plexes may be assumed to involve what is traditionally known as logical processes of deduction and induction, combined with background experience and knowledge of reality.

10. Concluding remarks

The proposals of Yngve (1996) for human linguistics pose problems for the understanding of how texts function in communications between authors and readers. In this paper it is presumed that methodologically much can be learned from traditional discourse analysis, as long as assumptions are not made about the independent existence of representations, formulations, or reconstructions of 'intended meanings', and as long as the focus is the understanding of the actual behaviour of 'communicating individuals' at particular times and locations. The lexical and textual features identified by authors such as Swales (1990), Hoey (1983, 1991, 1994), Winter (1977, 1994), Mann and Thompson (1988), and the contributors in Mann and Thompson (1992), should be seen not as 'carriers of meaning' or indicators of 'meaning-bearing' functions and relationships, but as links to internal states (plexes) of individuals. In practice, it may well be that the associations and networks identified by discourse analysts do not differ significantly from the associations and networks of individual plexes, and that, in other words, the analyses of 'intended meaning' and of underlying text macrostructures and superstructures are directly relevant to analyses of communicating individuals. It has certainly been the argument of this paper that there must be direct correspondences between observable textual elements and internal behavioural states and patterns of those that write and read texts, and that consequently the insights of 'traditional' discourse analysis are applicable within human linguistics.

11. References

Burazer, L. (2004): "Reconstituting notions of reference". In: Victor H.Yngve and Zdzisław Wąsik (eds.) *Hard-science linguistics* (London: Continuum, 2004), 115-128

Coulthard, M., ed. (1994): *Advances in written text analysis*. London: Routledge.

Daneš, F. (1974): "Functional sentence perspective and the organisation of text". In: Daneš, F. (ed.) *Papers in functional sentence perspective* (The Hague: Mouton), 106-128.

Firbas, J. (1966): "On defining the theme in functional sentence analysis", *Travaux Linguistique de Prague* 1, 267-280

Gundel, J. K. (1985): "'Shared knowledge' and topicality", *Journal of Pragmatics* 9(1), 83-107.

Halliday, M. A. K. and Hasan, Ruqaiya (1976): *Cohesion in English*. London: Longman.

Hoey, M. (1983): *On the surface of discourse*. London:Allen.

Hoey, M. (1991): *Patterns of lexis in text*. Oxford: Oxford Univ.Press.

Hoey, M. (1994): "Signalling in discourse: a functional analysis of a common discourse pattern in written and spoken English", in Coulthard (1994), 26-45.

Hutchins, W. J. (1977a): "On the structure of scientific texts", *UEA Papers in Linguistics* 5, 18-39

Hutchins, W. J. (1977b): "On the problem of 'aboutness' in document analysis", *Journal of Informatics* 1 (1), 17-35.

Hutchins, W.J. (2004): "Communicating scientific experiments in journal articles" In: Victor H.Yngve and Zdzisław Wąsik (eds.) *Hard-science linguistics* (London: Continuum, 2004), 304-317

Knorr-Cetina, K. D. (1981): *The manufacture of knowledge: an essay on the constructivist and contextual nature of science.* Oxford: Pergamon.

Mann, W.C. and Thompson, S.A. (1988): "Rhetorical structure theory: toward a functional theory of text organization", *Text* 8(3).243-281.

Mann, W.C. and Thompson, S.A. eds. (1992): *Discourse description: diverse linguistic analyses of a fund-raising text.* Amsterdam/Philadelphia: John Benjamins. (Pragmatics & Beyond, New series 16)

Meyer, G. (1992): "'In this paper we report …': speech acts and scientific facts", *Journal of Pragmatics* 17(4), 295-313.

Schank, R. C. (1975): *Conceptual information processing*. Amsterdam: North-Holland.

Sinclair, J.M. (1994): "Trust the text", in Coulthard (1994), 12-25.

Swales, J.M. (1990): *Genre analysis: English in academic and research settings.* Cambridge: Cambridge University Press.

Van Dijk, T. A. (1977): *Text and context: explorations in the semantics and pragmatics of discourse.* London: Longman.

Van Dijk, T.A. (1980): *Macrostructures: an interdisciplinary study of global structures in discourse, interaction, and cognition*. Hillsdale, NJ: Lawrence Erlbaum.

Winter, E. O. (1977): "A clause-relational approach to English texts: a study of some predictive lexical items in written discourse", *Instructional Science* 6, 1-92.

Winter, E. (1994): "Clause relations as information structure: two basic text structures in English", in Coulthard (1994), 46-68.

Yngve, V. H. (1996): *From grammar to science*. Amsterdam: Benjamins.