

TOWARDS A VALIDATED ANALYSIS OF SCIENTIFIC TEXT STRUCTURE

Graham Crookes

Introduction

The growth of science and the use of English as its principal language of information dissemination have vastly increased the number of science students using English as a second language (ESL) in institutions of higher learning around the world. In response to their needs (Julian, Lowenstein and Slattery 1979, Robertson 1983), we have seen the development of programs and research in English for Science and Technology (EST).

The rapid growth of EST in the 1970s resulted in a sudden demand for materials, and a variety of texts were published in a short space of time. These texts were designed to aid the development of skills necessary to handle types of scientific discourse. Their quality was variable. A characteristic deficiency was the absence of appropriate research or analysis of the nature of the tasks that students of science and technology had to carry out in English (Swales 1978, Mackay and Mountford 1978:vii), or of the materials, register and types of discourse they needed familiarity with. Such work has often displayed an uncoordinated and unsystematic approach to the issues investigated, and is often characterized by an inadequate data base or by a lack of rigor in the way data is gathered and analyzed. This undoubtedly reflects the fact that many practitioners work in situations where limited facilities for research are available (see e.g., Wingard 1981). Thus it has

usually been the case that investigators have proceeded on the assumption that explicit teaching of the characteristics of text will aid students ability to read and write such texts. The extent to which this assumption is valid for ESL is only now becoming clear, with the recent work of Carroll (1983, 1984). More seriously, little attempt has been made to support statements concerning the characteristics of ESP/EST text with hard evidence. Teachers and students have been left at the mercy of the materials writers' experience and intuition. A noted authority in the field has observed

ESP textbooks have been in many respects an educational failure.

(Swales 1980:11)

The quality and quantity of ESP/EST research has improved in recent years, as witnessed by the appearance of a new professional journal devoted to the field (The ESP Journal, begun in 1980), and by the formation of an ESP Special Interest Group in the TESOL organization. However, even the best of recent ESP research still has some of the limitations of earlier work.

The present work takes the scientific paper as its target genre. It considers existing ESP materials oriented analyses of the structure of the scientific paper, principally that of Swales (1981). Particular attention is focused on the need for validation of any analysis of the structure of a text type which is to be used as the basis for materials development. The latter part of the paper outlines one example of how this may be done.

The scientific paper

The scientific paper can be broadly defined as a type of

scientific writing, based on a single investigation, whose purpose is to contribute to the progress of science or technology (Peterson 1961:6). The present work is concerned only with papers published in professional, refereed, scientific journals. In addition, abstracts and "short papers", or communications, are excluded.

Morris (1966:204) identifies two basic types of scientific paper published in journals: the theoretical type, and the experimental-research paper. Peterson (1961:133, 169-170) refers to the review article as an additional important and separate type. The experimental-research paper, to quote Hill et al. (1982)

reports experimental or ex post facto research designed to test a hypothesis or theory.

(334)

According to the conventional wisdom of rhetoricians and technical writers, such papers can be assumed to incorporate different rhetorical structures. It is these structures which make them distinctive and which ESL/EST students have to be able to recognize if they are reading such papers, and produce, if they are writing them.

Scientific papers are subject to some constraints concerning form and style. The requirements of the journal editor and referees who represent the scientific community apply to all papers. The experimental-research paper must also reflect hypothetico-deductive scientific method. This leads to a degree of standardization which suggests that such papers may share a common basic structure or schema, or employ common units of discourse.

The need for EST materials concerned with the scientific paper

As Hill, Soppelsa and West (1982) observe, the ability to read and write experimental-research papers is important for success in the sciences. It has also been observed that science students exhibit a clear preference for technical, subject-related materials to be used in ESL classes. However, according to Schachter (1981), few ESL teachers are comfortable teaching students how to read and produce experimental-research papers. Until recently, there have been few if any ESL texts which deal with how to read and write such papers (Hill et al. 1982).

The pedagogical desirability of a knowledge of discourse structure has long been asserted. For example, Pitkin (1969) said, in relation to the teaching of English composition to native speakers:

We need presently and will continue to need more efficient models for teaching our students to read connected discourse with understanding.
(138)

Selinker, Todd Trimble and Trimble (1976) suggested that a similar need is felt in EST. Advanced ESL/EST students, they wrote

often seem unable to comprehend the total meaning of EST discourse even when they understand all of the words in each sentence and all of the sentences that make up the discourse.
(282)

For Selinker et al., it is the students' lack of knowledge of the rhetorical structure of the discourse which hinders them.

Both Hill et al. (1982), and Swales (1981) are concerned with similar problems in their work. These and other researchers assume a need for direct teaching of the rhetorical structure and

organization of scientific papers. It might be inferred from these concerns that knowledge of a text's structure aids comprehension and production of such texts. More recent research (Carrell 1983, 1984) supports this assumption. A relationship exists between an individual's psychological conception of a form and his ability to comprehend and utilize it. This 'conception' is widely referred to as a 'schema'.

Schema theory

Schema theory is a developing area of investigation in cognitive psychology. It is principally concerned with the way in which various types of background knowledge affect understanding and recall. Carrell (1983), in a recent review, distinguishes between formal schemata, which deal with the rhetorical structure of discourse, and content schemata, which deal with general world knowledge.

Schemata have been shown to guide the comprehension not only of events and actions...but also to guide the interpretation of the linguistic representation of these events, scenes, activities -- i.e., oral and written texts. (2)

Early work of relevance is Mandler and Johnson's (1978) work on story grammars. This refers to the underlying cognitive structure of a narrative, in terms of setting, event structure and episode. The proficient reader, Mandler and Johnson argue, has internalized such a story grammar, and uses it to process stories.

Further support for such a notion comes from the work of Kintsch and van Dijk (1975, 1978), which suggests that knowledge of the schema of a piece of discourse is critical for its

effective recall and understanding. In Kintsch and van Dijk (1978) they further develop a model of text processing, and partially summarize their earlier work:

There are a number of highly conventionalized text types. If a reader processes such texts in accordance with their conventional nature, specified well-defined schemata are obtained. These are shared by the members of a given cultural group and, hence, are highly suitable for research purposes. Familiar examples of such texts are stories (Kintsch and van Dijk 1975) and psychology research reports (Kintsch 1974 [17-22]). These schemata specify both the schematic categories of the texts (e.g. a research report is supposed to contain introduction, method, results and discussion sections) as well as what information in each section is relevant to the macrostructure (e.g., the introduction of a research report must specify the purpose of the study)

(373)

For Kintsch and van Dijk, it is the schema which determines which of the many propositions in a text are relevant or irrelevant to the reader, and thus directly affects how and whether they are processed or recalled. If the reader has a

schema...that is not well defined, the outcome of the processing of the text will be "haphazard"

(375)

with obvious problems for comprehension, whereas if s/he is familiar with the conventional nature of the text, well-defined schemata will be produced which will aid comprehension and recall. The reference to the culture-specific nature of schema is borne out, at least for the narrative genre, by the work of Colby (1970). The ESL/EST student who does not share such schematas may experience difficulties in reading comprehension.

The work of Kintsch and van Dijk demonstrates how knowledge of schema or text structure is reflected in comprehension. It

provides a justification for future analyses of text structure. It also provides needed support for earlier investigations into text structure which used different and possibly weaker methodologies, such as Lackstrom, Selinker and Trimble (1973), Hepworth (1979), Hutchins (1977), Hoey (1979), Day (1982) and others.

Inadequacy of existing analyses

The main problem with existing analyses arises out of the characteristics of the rhetoric tradition, which forms the basis for such work. This tradition prescribes, but does not attempt to verify that what it prescribes is what happens.

Day's work, for example, although thorough and useful, cannot be taken as establishing the accuracy of the system advocated by technical writers. It is prescriptive, and therefore not necessarily descriptively adequate. The fact that composition manuals and writing teacher advocate the use of certain rhetorical structures in technical writing does not necessarily mean that such structures are in fact employed in technical writing.

Similar problems attend the work of Selinker and his colleagues. As Calfee and Curley (1984:168) observe, while it warrants attention, it is purely descriptive, and no attempt is made to support their hypothesized structure with experimental findings.

Alternatively, in the case of Hutchins, (1977) the background to the work is principally stylistics. As a result, the analysis is deliberately abstract, intended to provide a way of perceiving a system at some underlying level. Besides the

difficulties this poses, Hutchins' conclusions concerning the structure of the scientific paper are weakened by his heavy reliance on Gopnik's Linguistic Structures in Scientific Texts (1972). He assumes that Gopnik's conclusions relate to the scientific paper, when her data base was in fact the preconference abstracts of presentations to be given at a scientific meeting.

The principal difficulties with Hepworth's (1979) work are simple methodological problems. To touch very briefly on main points: the writer does not give any indication what the corpus he analysed was, or how it was selected, so one has no way of knowing how far generalizations made on its basis can be taken. In addition, no apology is made for a reliance on intuition in what is presumably supposed to be a scientific investigation (Hepworth 1979:155).

It may be argued that we cannot assign anything more to these analyses than a verdict of 'not proven', since their units (as Hoey 1979 points out) are not clearly defined, their analytical techniques, when used, are not indicated, nor are their corpuses referenced (in the case of Hepworth) or appropriate (in the case of Hutchins). Above all, no attempt is made to subject them to empirical test. However, it should be mentioned that the general line of argument in all cases is similar. In particular, there is quite close agreement between the structure arrived at by Hutchins, and that of Swales.

The work of Swales is far more rigorous than earlier work. With a more satisfactory data base (16 articles from each of

physics, biology/medicine, and the social sciences), Swales identified four 'moves' in the introductory sections of the papers, which in turn could be sub-categorized.

- Move 1 Establishing the field
 - [by] showing centrality
 - stating current knowledge
 - ascribing key characteristics
 - Move 2 Summarizing previous research
 - Move 3 Preparing for present research
 - Move 4 Introducing present research
 - [by] giving the purpose
 - describing present research
- (22a)

Here, for the first time, we are presented with a referenced and appropriate corpus, units identified, a carefully detailed explanation of the analysis of the corpus and of remaining problems. There is only one major problem - the research lacks empirical validation. Swales himself indicates his awareness of this. He recognizes the danger that

the discourse analyst labels something as x
and then begins to see x occurring all over
the place

(Swales 1981:13)

"One way out" that he suggests, which various EST analysts have taken (Tarone, Dwyer, Gillette and Icke 1981, Pettinari 1981 and others) is to consult a specialist as to what is 'really' going on. Swales did not do this. He recognizes that

I am open to the charge that
my unsubstantiated and ill-defined
terminological labels...are little more than
a reflection of my own perceptual
predispositions.

(14)

Part of the object of the present work is to indicate how an analysis of this sort may avoid laying itself open to such a charge.

Other difficulties which exist apply to the sampling. There are two valid approaches: either one should follow Roe (1977), and select according to the texts used by a particular group of EST students or courses, or a random sample could be taken and the procedure indicated. Swales' work grew out of an investigation of the structure of the section of article introductions which reports previous research. This limits it, in that although he did not confine his work to this area, the original orientation of the work resulted in problems concerning the corpus selected. Full details of the random selection method used are not given, but it is clear that Swales deliberately selected only those articles which contain reports of previous research within them. Having widened the investigation to the general structure of article introductions, his conclusions are limited in their generalizability by this somewhat inappropriate corpus. He does mention this (19) in terms of the four-part analysis he proposes "of which part 2 would occur a priori", but not in terms of the effect this would appear to have on the generalizability of the conclusions. The basic conclusion is that, generally, authors make four 'moves' in article introductions: one of each of the four mentioned, and in numerical order.

Validating a discourse analysis

In developing an analysis of behavior which proposes that a particular type of behavior is made up of sequences of units, it is standard practice for the accuracy of such an analysis to be established by showing that the units can be defined in such a way that a group of trained raters can record the incidence of

units of behavior at a sufficiently high level of agreement. If a system represents a realistic depiction of the activity, it will be possible to obtain such agreement. The Swales model is open to the criticism that it is a purely subjective analysis, as Swales himself recognises. But if it can be shown that it can be applied by a group of trained raters to an appropriate corpus at an adequate level of agreement, a strong defence is made against this criticism. This was what the present study attempted to do.

Method

In outline, the procedure was as follows:

- 1) a corpus was selected
- 2) raters were selected
- 3) the overall design was explained to the raters
- 4) definitions of the units of analysis were presented and discussed
- 5) unit boundary markers were presented and discussed
- 6) worked examples were presented
- 7) raters practiced analysing simple texts, whose structure had already been established
- 8) more complex texts were analysed, interrater reliability scores calculated, and disagreements discussed by the raters, both between themselves and with the trainer
- 9) step 8 was repeated until a satisfactory level of interrater agreement was attained.
- 10) the corpus was rated
- 11) analysis

The corpus

Initially, a corpus of 96 scientific articles was selected. The same basic categories as used by Swales were adopted, for reasons of comparability: "the hard sciences", "'the Biology/Medical' field" and "the 'Social Sciences'". Within each of these three sections, four journals were chosen, according to Garfield's criterion of popularity. Garfield (1981, 1982) ranks journals according to number of times cited in a particular period. Thus it is possible to determine the importance of a particular journal in terms of its likelihood of being encountered by anyone reading or doing research in a given area. By working down the unified rank list for times cited in the post-1980 period, it was possible to build up a group of the four most "popular" journals for each of the three areas.

Having settled upon the titles, and having chosen 1980 and after as the period that selection would be restricted to, a table of random numbers was used to provide a point of entry to a particular month in the period from 1/80. From the issue thus selected (if the journal appeared monthly) and the first of the month otherwise, four articles were then selected. Articles were examined, beginning with the first in the relevant issue, and in sequence thereafter, and the first four suitable were entered into the corpus. Articles were rejected if they appeared to be written exclusively by non-native authors, if they appeared to be review or theoretical articles, and if selection would mean having more than one article by the same author in the corpus. If four suitable articles were not to be found in the same issue of the journal, the subsequent issue was moved to, and so on.

Training of raters; testing the Swales model

It was decided that raters should be individuals with some linguistic sophistication. Specialists (i.e. scientists or technical writers) were not selected, principally due to the logistical problems envisaged.

A pair of MA(ESL) students was trained (the writer and one other) using the extensive sequence of fully analysed article introductions and representative examples given by Swales, as well as a number of article introductions not included in the corpus, but taken from the same journals.

One quarter of the corpus was selected (by stratified random sampling from the original corpus), and rated. Despite ten hours of joint effort, plus individual study of the materials, satisfactory interrater agreement could not be arrived at, unless articles which appeared not to conform to the Swales model were excluded. For the remainder, interrater reliability was high: $\kappa = 0.96$ (Cohen 1960).

Difficulty with the system did not merely reflect the raters' lack of training, intelligence, or understanding of English - it rapidly became clear that some introductions deviated so far from a strict four-move schema as to call into question such a system's descriptive adequacy. The terms used in the Swales model to characterize moves are brief. Their explanations are very detailed, but despite this some appeared to be too vague to apply without unacceptable levels of disagreement.

More significant was the question of boundary markers.

Brown and Yule (1983:69) refer to formulaic expressions which are used as markers to indicate the boundaries of blocks of discourse. When there are no boundary markers, reliance must be placed on the notion of topic, which, they say, is very difficult to define. Swales, referring to Move 2, says "in contrast to the other three moves, there tends to be no signal of onset." This was generally found to be the case in the present work. Swales also gives ten examples of the openings of Move 2s, which, unfortunately, as he says, "exhibit an uncanny resemblance to Move 1 signals" (33-34). An explanation of the difference between Move 1 and Move 2 signals is given (34), but the problem caused by this similarity in signals was too severe to be overcome. Authors exhibited a disturbing tendency to summarize previous work (the definition of Move 2) at any place in the Introduction. It became clear that revision of both Move 1 and Move 2 was necessary to clarify the point at which the former changed into the latter.

The other major problem concerned the applicability of the conception that, fundamentally, Introductions have four moves, one of each type. Swales' corpus does not appear to contain any introductions with more than four moves, except in one exceptional case. There, "the first of two Move 3 elements is embedded inside Move 2." If there is not a hierarchical relationship implied by the word 'embedding' (impermissible given the definition of the moves), then the alternative is a Move 2, followed by a Move 3, followed by a Move 2. It seemed that, given our corpus and the definitions derived from Swales' terms and explanations, there were very many examples of such

'embedding', involving all four moves.

It appeared to be the case that though the units of analysis were well motivated (despite a possible Move 5 -- see Section IV), their definitions were insufficiently clear in some cases. The limitations on the sequences in which they might appear and their optionality implied by the Swales model did not permit its easy application to the corpus at hand.

Consequently, a slightly revised model was developed. For the revised version, short definitions were prepared. Units of discourse were referred to as 'types' rather than 'moves'. An emphasis on conciseness, and on changes in level of generality was introduced with regard to Type 1s. The function of reporting past research for the purpose of summarizing it was introduced for Type 2s. The definitions of the other two units reflect more closely the original Swales units. The requirements that there be only four units, and that they be in the order 1-2-3-4, were both eliminated. All introductions were to be coded - there was no category for 'deviants'. Finally, three new raters, again all MA(ESL) students were obtained. Following training, raters attained an interrater-agreement figure of 0.9 (Fliess 1971) on a test run of a small section of the first quarter of the corpus. They then rated completely the second quarter of the corpus. Financial and temporal limitations precluded rating the whole of the corpus: a further one quarter of the original corpus was coded (different to that used to test the Swales model). The results are summarized in Table 1.

INSERT TABLE ONE ABOUT HERE

Results

As may be seen from the table, in three-quarters of the sample, raters were able to reach agreement as measured by kappa (Fliess 1971) at above 0.6 (see also Gelfand and Hartman 1975, below). There were two texts for each journal. Sentences which were not coded unanimously were omitted in determining the sequence of units in each text. No structure is indicated for texts where kappa was below 0.6. The most common structures were 24 and 1234 (five occurrences each). The structure 1234 was not observed in any social science texts. The structure 24 was observed once in social science texts. There were on average 1.8 times more units in social science texts than in other texts.

Discussion

It may be tentatively concluded that results are consistent with the idea that four basic units of discourse occur in scientific experimental-article introductions. In some shorter article introductions, there is a tendency to find the simple four-move schema posited by Swales, but this is by no means the only possibility. In more complex passages, a variety of alternative arrangements are possible, involving repeated use of mainly the 'internal' units Two and Three. In the corpus under consideration, writers normally, though not universally, begin with a One and end with a Four (11/16 times).

Results must be interpreted with caution, since the figures for interrater agreement are low. Hartman (1977) comments

No entirely agreed upon set of rules for deciding upon an acceptable value for trial (or session) reliability has yet been formulated.

(113-4)

He quotes Gelfand and Hartman (1975) as "recommending that...kappa...should exceed 0.6". The statistic, used in the present study, is Fliess' (1971) generalisation of kappa to more than two raters. It is a non-parametric statistic derived from nominal categorial data. It is the most conservative of the measures available (others being pairwise kappa, phi, and percentage agreement). In the present study, generalizations concerning introduction structure are based upon the 75% of cases where kappa is greater than 0.6.

The unit of coding was the sentence. The sentence was selected as the basic unit of analysis since it initially appeared that writers reflected the traditional conception of the sentence as constituting a complete unit of meaning (or 'thought'). Difficulties were envisaged in choosing a smaller but perhaps less well-defined unit, such as the clause or phrase. However, in some cases, writers utilized the flexibility of language, and complex sentences, to produce a sentence which contained elements of more than one unit of text, according to the system's definitions. Thus

Interest in these reactions has been sparked by the hope that clusters might possess unique catalytic activity, and by the proposal [reference] that clusters might be used as soluble models for metal surfaces.

Bavaro, Montangero & Keister (1983:4977)

In this case, following a previous Type One, the writer continues to refer to the interest of the topic (a major Type One lexical signal) but also references past work and raises two hypotheses. The raters in this case were obliged to decide which of these characteristics was most outstanding - were the writers primarily

indicating the interest of the topic, or were they raising hypotheses which had already been mentioned in the literature? This case, and others like it, were not always coded unanimously. Another particularly clear example of disagreement was the following:

Although fluorescent antibody studies have established the overall distribution of myosin within some vertebrate nonmuscle cells (1,2,8,11,19,20,21-24), the form and detailed arrangement of myosin has remained a mystery.
(Herman & Pollard 1981:346)

This sentence might have been written

Fluorescent antibody studies have established the overall distribution of myosin within some vertebrate nonmuscle cells (1,2,8,11,19,20,21-24). However, the form and detailed arrangement of myosin has remained a mystery.

Had this been the case, unanimous agreement might have been expected, to the effect that the first sentence constituted a Type 2, and the second a Type 3. By taking the sentence as the basic unit of coding and forcing a decision, this fuzziness of boundaries necessarily manifested itself in rater disagreement at such points.

Other sources of disagreement are simply what appear to be mistakes, where raters disagreed over what appears to this writer to be a clearcut case. This must be taken as evidence of raters' lack of adequate training. The variability of texts suggest also, that a longer test run should have been undertaken. The effects of raters' understanding of topic are also a possible source of error.

Brown and Yule (1983) make instructive comments about the relationship between "formulaic expressions such as 'Once upon a

time" and topic change as boundary markers. They point out that topic is something which is difficult to define and perceive clearly. Brown and Yule refer to it as being located in the writer rather than in the text. Genres which make heavy use of formulaic expressions are better suited to the present type of analysis. Attempting the present form of analysis outside of such genres as the fairy story or the scientific experimental-research paper may be inadvisable.

At the outset of the inquiry, it was not clear what the balance was between topic change and formulaic phrase in determining the divisions between units of text in scientific experimental-research article introductions. Although the genre is formalised, variation exists in the degree to which writers adhere to such formalism. In particular, writers vary in the extent to which they use formulaic expressions to mark transition from one block to another, as opposed to relying merely on topic change. They may be following a simple, standard pattern, but if this is not signaled explicitly, it is left to the discourse analyst's or rater's understanding of the topic as a whole, and relations between different aspects of it, to decide where one block finishes and another starts. Thus, to give a concise example

...These methods are useful for comparing intact filaments in different tissues. They do not indicate whether individual filament forming polypeptides (FFP's) have significantly large regions of related aminoacid sequences.

(Milstone and McGuire 1981:312)

In this case, even a close reading by a non-specialist may miss the fact that the second sentence is raising a gap. It is

necessary to have an understanding of the relationship between aminoacid sequences and FFPs. Such an understanding is naturally taken for granted by the writers. There is also the negative in the second sentence, but this on its own is dangerously little to go on in coding the passage. There is no 'but', no 'a weakness is...', nor any of the large number of other possible lexical signals the writers could have used. A heavy reliance is placed on topic alone.

In the present work, a decision was made to use only non-specialists, principally for practical reasons. Also, even if it had been possible to gather a group of, say, astrophysics experts and have them rate astrophysics papers, conclusions about the structure of the general scientific article introduction would have been impossible. It is necessary to point out that the raters' lack of full comprehension of the topic of some of the articles may have limited the accuracy of their codings. However, it may be argued that although levels of interrater agreement are not uniformly high, this does not invalidate generalisations made concerning text structure drawn here. In addition, as de Beaugrande and Dressler (1981) point out

A science of texts demands its own terms and notions because of the nature of its object...we should work to discover regularities...rather than rules or laws....It is the task of science to systemize the fuzziness of its objects of inquiry, not to ignore it or argue it away.

(xiv-xv)

Finally, since it might be expected that any future analysis of this sort would be carried out by materials writers, the development of an analysis capable of being applied by non-specialists is desirable.

Subject-specific variation

There is some suggestion that the incidence of repeated cycles of internal moves increases according to subject: articles in social science journals tended to have much longer and more complex introductions. There were 1.8 times more units in social science introductions as in the other two groupings. Perhaps the newness of the field results in a lack of shared preconceptions and a greater need for both definition of terms and motivation of hypotheses. Considering also articles in the social sciences from outside the corpus itself (but from the same journals and type of articles), there was occasional use of topic-specific subheadings within the introduction, which did not occur in other journals. More significantly, stretches of unreferenced text presenting theoretical background sometimes occurred. Although there were none in the quarter of the corpus on which rating was done, their existence suggests the possibility of a fifth unit, whose function would be "presenting general, non-referenced theoretical background".

Articles which were not primarily experimental or data-based were not included in the corpus. It should be obvious that articles on pure mathematics, for example, are a type which the current analysis would not apply to, since its units presuppose data-based work in the hypothetico-deductive mode of science. Articles of the social sciences which take mathematics as a model -- economics, for example, and some parts of linguistics -- would equally be unsuitable for the present analytical framework.

Pedagogical and materials-writing implications

Pedagogical implications which may be drawn from this study apply more to materials designed to aid EST reading than writing, because of the product-based nature of the analysis. Such materials would, as Swales has suggested, be aimed at making salient the possible structures of article introductions, devices used to signal them, and the way relations between the topics of different segments contribute to the text as a whole. However, the main point to be made is that the analysis such materials are to be based on must be a valid reflection of discourse structure in target texts, and proven so. If we inculcate an inaccurate schema into our ESL/EST students concerning the expectations they should have of a text, we are making their task harder, not easier.

Further research

When the present work was initiated, it was envisaged as merely verifying existing analyses of the 'Introduction', and then going on to deal with the 'Discussion'. Belanger (personal communication) has begun work on the structure of the 'Discussion' section, which is of course, the natural progression from the present work. Indications are that a structure consisting of a number of basic elements which may be repeated in various subsequences is to be found in the 'Discussion' section, as in the 'Introduction'. It is anticipated that verification of such an analysis using the present technique may be more difficult, because the structure of the discussion section is believed to be less conventionalized and harder for even technical writers to make prescriptions about (Day 1982, see

above). Because of the probable greater importance of topic in this section, it may be impractical to attempt an analysis across sciences. Instead, a subject-specific or journal-specific analysis may need to be undertaken, using a group of expert raters.

Finally, the work presented here has focused solely on product. Research in this area is also needed concerning process, as Swales (personal communication) has suggested. Other techniques besides the use of appropriate schema are used by the good reader. In the present context, a knowledge of how the good ESL/EST reader handles scientific texts would provide information concerning the processes involved in reading. More broadly, research is also needed concerning how such texts are generated. Suitable techniques for research on composing and revising processes already exist (see Heuring 1984 for a recent review). Findings arrived at from this perspective would show the other side of the subject under investigation, and would have implications for the writing of EST materials which could complement those concerning EST reading materials.

Notes

- 1 The present article is derived from the writer's MA thesis, 'Towards a validated discourse analysis of scientific text'. The support and advice of Professors J.C. Richards, M.H. Long and C. Chaudron is gratefully acknowledged.
- 2 A complete listing of the corpus used is available on request.

TABLE ONE

journal	text no.	structure (text units)	interrater agreement (kappa)	length (sentences)
<u>hard sciences</u>				
Astrophys.J.	1	1 2 4 2 4	0.61	17
" "	2		0.49	10
Phys.Rev.B.	1	2 4	0.62	9
" " "	2	1 2 3 4	1.00	6
J.Chem.Phys.	1	1 2 3 4	0.66	23
" " "	2	1 2 3 4	0.93	19
J.Am.Chem.Soc.	1		0.59	14
" " " "	2	1 2 4	0.78	17
=====				
<u>biology/medical sciences</u>				
New Eng.J.Med.	1	1 2 3 4	0.91	12
" " " "	2		0.50	9
Lancet	1	2 4	0.77	6
" "	2		0.55	10
J.Cell Biology	1		0.33	7
" " "	2	1 2 3 4	0.74	11
J.Physiology	1	2 3 4	0.70	11
" "	2	2 4	0.70	8
=====				
<u>social sciences</u>				
Child Dev.	1	1 2 3 2 3 4	0.72	36
" "	2	1 2 3 2 3 2 3	0.61	31
Arch.Gen.Psych.	1	1 2 3 2 3 4	0.76	10
" " "	2	1 2 4 2 4	0.65	14
Am.Soc.Rev.	1	2 4	1.00	17
" " "	2	1 2 3 2 3 2 3 4	0.83	39
J.Pers.Soc.Psych.	1		0.24	74
" " " "	2	2 4 2 3 2 3 4 2 4	0.68	57
=====				

- Allen, J.P.B. & H.G. Widdowson. 1978. Teaching the communicative use of English. In Mackay & Mountford 1978.
- Bates, M. 1978. Writing 'Nucleus'. In Mackay & Mountford 1978.
- Bavaro, L.M., P. Montangero & J.B. Keister. 1983. Kinetics and mechanisms of oxidative addition and reductive elimination of hydrogen in triruthenium clusters. Journal of the American Chemical Society 105:15. 4977-4981.
- Beaugrande, R. de & W. Dressler. 1981. Introduction to text linguistics. London: Longmans.
- Belanger, M. 1984. A preliminary analysis of the structure of the discussion sections in ten neuroscience journal articles. Mimeo.
- Brown, G. & G. Yule. 1983. Discourse analysis. Cambridge: CUP.
- Calfee, R.C. & R. Curley. 1984. Structures of prose in the content areas. In J. Flood (ed.), Understanding reading comprehension. Newark, Delaware: International Reading Association.
- Carrell, P.L. 1983. Some issues in studying the role of schemata, or background knowledge, in second language comprehension. Paper presented at the 1983 TESOL Convention, Toronto, Ontario. March 16, 1983.
- _____. 1984. Facilitating reading comprehension by teaching text structure: what the research shows. Paper presented at the 18th Annual TESOL Convention, Houston, Texas, March 1984.
- Cohen, J. 1960. A coefficient of agreement for nominal scales. Educational and Psychological Measurement 20. 37-46.
- Colby, B.N. 1970. The description of narrative structures. In (ed.) P.L. Garvin, Cognition: a multiple view. N.Y.: Spartan.
- Cohen, J.A. 1960. A coefficient of agreement for nominal scales. Educational and Psychological Measurement 48:1. 157-184
- Day, R.A. 1979. How to write and publish a scientific paper. Philadelphia: ISI.
- Fliess, J.L. 1971. Measuring nominal scale agreement among many raters. Psychological Bulletin 76:5. 378-382.
- Garfield, E. (ed.) 1981. Science Citation Index. Philadelphia: ISI.
- _____. 1982. Social Science Citation Index. Philadelphia: ISI.

- Gelfand, D.M. and Hartmann, D.P. 1975. Child behavior analysis and therapy. N.Y.: Pergamon.
- Gopnik, M. 1972. Linguistic structures in scientific texts. Hague: Mouton.
- Hartmann, D.P. 1977. Considerations in the choice of interobserver reliability estimates. Journal of Applied Behavior Analysis 10:1. 103-116.
- Hepworth, G.R. 1979. Rhetorical competence and EST discourse. In C. Yorio et al. (eds.), On TESOL '79. Washington, DC: TESOL.
- Herman, I.M. & T.D. Pollard. 1981. Electron microscopic localization of cytoplasmic myosin with ferritin-labeled antibodies. Journal of Cell Biology 88. 346-351.
- Heuring, D. 1984. Revision strategies of skilled and unskilled writers: five case studies. M.A. Thesis, University of Hawaii.
- Hill, S.S., B.F. Soppelsa and G.K. West. 1982. Teaching ESL students to read and write experimental-research papers. TESOL Quarterly 16:3.
- Hoey, M.P. 1979. Signalling in discourse. Birmingham, UK: English Language Research Unit, University of Birmingham.
- Hutchins, J. 1977. On the structure of scientific text. University of East Anglia Papers in Linguistics 5.
- Julian, A.C., J. Lowenstein & R.E. Slattery (eds.). 1978. Open doors: 1977-78. N.Y.: Institute for International Education.
- Kintsch, W. 1974. The representation of meaning in memory. N.Y.: Erlbaum.
- Kintsch, W. & van Dijk, T.A. 1975. Comment on se rapelle et on r sume des histoires. Langages 40. 98-116.
- Kintsch, W. & van Dijk, T.A. 1978. Towards a model of text comprehension and production. Psychological Review 85. 363-394.
- Lackstrom, J., L. Selinker & L. Trimble. 1970. Grammar and technical English. In R.C. Lugton (ed.), English as a second language: current issues. Philadelphia, Pa.: Center for Curriculum Development.
- _____. 1973. Technical rhetorical principles and grammatical choice. TESOL Quarterly 7:2. 127-136.
- Mackay, R. & A. Mountford (eds.). 1978. English for Special Purposes: a case study approach. London: Longman.
- Mandler, J.M. & Johnson, N.S. 1977. Remembrance of things past:

story structure and recall. Cognitive Psychology 9. 111-151.

Milstone, L.M. & J. McGuire. 1981. Different polypeptides form the intermediate filaments in bovine hoof and esophageal epithelium and in aortic endothelium. Journal of Cell Biology 88. 312-316.

Morris, J.E. 1966. Principles of scientific and technical writing. N.Y.: McGraw-Hill.

O'Connor, M. 1978. Editing scientific books and journals. London: Pitman Medical.

Peterson, M.S. 1961. Scientific thinking and scientific writing. NY: Reinhold.

Pettinari, C. 1981. The function of a grammatical alternation in 14 surgical reports. Mimeo, University of Michigan.

Pitkin, W.J. 1969. Discourse blocs. College composition and communication 20:2.

Robertson, D.L. 1983. English use, needs and proficiency among foreign students at the University of Illinois at Urbana-Champaign. PhD (Dept. of Edn.) University of Illinois.

Roe, P. 1977. Scientific text. Birmingham, UK: ELR.

Schachter, J. 1981. Teach EST - me? No thanks. Paper presented at the 15th Annual International Conference of TESOL, Detroit MI, March 1981.

Selinker, L., R.M. Todd Trimble and L. Trimble. 1976. Presuppositional rhetorical information in EST discourse. TESOL Quarterly 10:3.

_____. (eds.) 1981. English for academic and technical purposes. Rowley: Newbury House.

Swales, J. 1980. ESP: the textbook problem. ESP Journal 1:1. 11-25.

_____. 1981. Aspects of article introductions. Aston ESP Research Report #1. Birmingham, England: University of Aston.

Tarone, E., S. Dwyer, S. Gillette & V. Icke. 1981. On the use of the passive in two astrophysics journal papers. The ESP Journal 1:2.

Wingard, P. 1981. Some verb forms and functions in six medical texts. In L. Selinker, E. Tarone & V. Hanzeli (eds.), English for Academic and Technical Purposes. Rowley, Mass.: Newbury House.