Towards a Validated Analysis of Scientific Text Structure

GRAHAM CROOKES

University of Hawaii

1. 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 was 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 1980, Mackay and Mountford 1978: vii), or of the materials, register, and types of discourse they needed to be familiar 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 are gathered and analysed. 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. However, 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. It has been observed that ‘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-orientated 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.

2. 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, 169-70) 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' (p. 344).

According to the conventional wisdom of rhetoricians and technical writers, different types of paper 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 experimental-research papers may share a common basic structure or schema, or employ common units of discourse.

3. 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 when 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 scientific papers of this type (Hill et al. 1982).

The pedagogical desirability of a knowledge of discourse structure has long been asserted. For example, Pitkin (1969: 138) observed, 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.

Selinker, Todd Trimble, and Trimble (1976: 282) 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.

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 or her ability to comprehend and utilize it. This ‘conception’ is widely referred to as a ‘schema’.

4. 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: 2), 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.

Early work of relevance is Mandler and Johnson’s (1977) 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: 373) 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).

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”’ (p. 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. Kintsch and van Dijk’s reference (1978: 373) to schemata being ‘shared by the members of a given cultural group’ (i.e. culture-specific) is borne out, at least for the narrative genre, by the work of Colby (1970). The ESL/EST student who does not share such schemata 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 other 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 (1979), and others.

5. 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 teachers 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 analysis is deliberately abstract, intended to provide a way of perceiving a system at some underlying level. Besides the difficulties this poses, Hutchins’s 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 pre-conference abstracts of presentation 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.

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 corpus 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 (sixteen articles from each of physics, biology/medicine, and the social sciences), Swales (1981: 22a) 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.
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’ (1981: 13). ‘One way out’ that he suggests, which various EST analysts have taken (Tarone, Dwyer, Gillette, and Icke 1981, Pettinari 1983, 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. (p. 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’s 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. His investigation having been widened to the general structure of article introductions, his conclusions are limited in their generalizability by this somewhat inappropriate corpus. He does mention this (p. 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.

6. 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. This procedure may similarly be applied to an analysis of text. The Swales model is open to the criticism that, however explicit its criteria and its exemplification, it remains in the end based on personal and individual judgement, as Swales himself recognizes. 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.

6.1. 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 practised analysing simple texts, whose structure had already been established.
8 More complex texts were analysed, inter-rater reliability scores calculated, and disagreements discussed by the raters, both among themselves and with the trainer.
9 Step 8 was repeated, until a satisfactory level of inter-rater agreement was attained.
10 The corpus was rated.
11 Analysis.

6.2 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 to which selection would be restricted, a table of random numbers was used to provide a point of entry to a particular month in the period from January 1980. 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 found in the same issue of the journal, the subsequent issue was moved to, and so on.

6.3 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 inter-rater agreement could not be arrived at, unless
articles which appeared not to conform to the Swales model were excluded. For the remainder, inter-rater 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’ (33-4). 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’s 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’s terms and explanations, there were very many examples of such ‘embedding’, involving all four moves.

It appeared to be the case that although the units of analysis were well motivated (despite a possible Move 5—see below), 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 inter-rater agreement figure of 0.9 (Flyss 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 from that used to test the Swales model). The results are summarized in Table 1.

<table>
<thead>
<tr>
<th>Journal</th>
<th>Text No.</th>
<th>Structure (text units)</th>
<th>Inter-rater agreement (kappa)</th>
<th>Length (sentences)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hard sciences</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Astrophysical Journal</td>
<td>1</td>
<td>1 2 4 2 4</td>
<td>0.61</td>
<td>17</td>
</tr>
<tr>
<td>Astrophysical Journal</td>
<td>2</td>
<td></td>
<td>0.49</td>
<td>10</td>
</tr>
<tr>
<td>Physical Review Part B</td>
<td>1</td>
<td>2 4</td>
<td>0.62</td>
<td>9</td>
</tr>
<tr>
<td>Physical Review Part B</td>
<td>2</td>
<td>1 2 3 4</td>
<td>1.00</td>
<td>6</td>
</tr>
<tr>
<td>Journal of Chemical Physics</td>
<td>1</td>
<td>1 2 3 4</td>
<td>0.66</td>
<td>23</td>
</tr>
<tr>
<td>Journal of Chemical Physics</td>
<td>2</td>
<td>1 2 3 4</td>
<td>0.93</td>
<td>19</td>
</tr>
<tr>
<td>Journal of the American Chemical Society</td>
<td>1</td>
<td></td>
<td>0.59</td>
<td>14</td>
</tr>
<tr>
<td>Journal of the American Chemical Society</td>
<td>2</td>
<td>1 2 4</td>
<td>0.78</td>
<td>17</td>
</tr>
<tr>
<td><strong>Biology / medical sciences</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>New England Journal of Medicine</td>
<td>1</td>
<td>1 2 3 4</td>
<td>0.91</td>
<td>12</td>
</tr>
<tr>
<td>New England Journal of Medicine</td>
<td>2</td>
<td></td>
<td>0.50</td>
<td>9</td>
</tr>
<tr>
<td>The Lancet</td>
<td>1</td>
<td>2 4</td>
<td>0.77</td>
<td>6</td>
</tr>
<tr>
<td>The Lancet</td>
<td>2</td>
<td></td>
<td>0.55</td>
<td>10</td>
</tr>
<tr>
<td>Journal of Cell Biology</td>
<td>1</td>
<td></td>
<td>0.33</td>
<td>7</td>
</tr>
<tr>
<td>Journal of Cell Biology</td>
<td>2</td>
<td>1 2 3 4</td>
<td>0.74</td>
<td>11</td>
</tr>
<tr>
<td>Journal of Physiology</td>
<td>1</td>
<td>2 3 4</td>
<td>0.70</td>
<td>11</td>
</tr>
<tr>
<td>Journal of Physiology</td>
<td>2</td>
<td>2 4</td>
<td>0.70</td>
<td>8</td>
</tr>
<tr>
<td><strong>Social Sciences</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Child Development</td>
<td>1</td>
<td>1 2 3 2 3 4</td>
<td>0.72</td>
<td>36</td>
</tr>
<tr>
<td>Child Development</td>
<td>2</td>
<td>1 2 3 2 3 2 3</td>
<td>0.61</td>
<td>31</td>
</tr>
<tr>
<td>Archives of General Psychiatry</td>
<td>1</td>
<td>1 2 3 2 3 4</td>
<td>0.76</td>
<td>10</td>
</tr>
<tr>
<td>Archives of General Psychiatry</td>
<td>2</td>
<td>1 2 4 2 4</td>
<td>0.65</td>
<td>14</td>
</tr>
<tr>
<td>American Sociological Review</td>
<td>1</td>
<td>2 4</td>
<td>1.00</td>
<td>17</td>
</tr>
<tr>
<td>American Sociological Review</td>
<td>2</td>
<td>1 2 3 2 3 2 3 4</td>
<td>0.83</td>
<td>39</td>
</tr>
<tr>
<td>Journal of Personality and Social Psychology</td>
<td>1</td>
<td></td>
<td>0.24</td>
<td>74</td>
</tr>
<tr>
<td>Journal of Personality and Social Psychology</td>
<td>2</td>
<td>2 4 2 3 2 3 4 2 4</td>
<td>0.68</td>
<td>57</td>
</tr>
</tbody>
</table>
6.4 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 2-4 and 1-2-3-4 (five occurrences each). The structure 1-2-3-4 was not observed in any social science texts. The structure 2-4 was observed once in social science texts. There were on average 1.8 times more units in social science texts than in other texts.

7. 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 is 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 inter-rater agreement are low. Hartman (1977: 113-14) comments

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

He quotes Gelfand and Hartman (1975) as ‘recommending that . . . kappa . . . should exceed 0.6’. The statistic used in the present study is Fliess’s (1971) generalization 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 per cent 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, and 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 and 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 a 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 suggests 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 and topic change as boundary markers. They point out that topic is something which is difficult to define and perceive clearly. Scientific papers of the type under discussion make use of standard formulas and phrases (and variations on them) in many places. At the outset of the inquiry, it was not clear what the balance was between topic change and such standard phrases in determining the divisions between units of text in scientific experimental-research article introductions. Although the genre is somewhat formalized, variation exists in the degree to which writers make use of such devices in marking transitions 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 signalled 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 (FFPs) 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 that 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 inter-rater agreement are not uniformly high, this does not invalidate generalizations made concerning text structure drawn here. In addition, as de Beaugrande and Dressler (1981: xiv–xv) 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.

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.

7.1 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 than 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 hypothesis. 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 un-referenced 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.

7.2 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 that 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.

7.3 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 (1984) 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 1979, 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. (Received December 1984)

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, and the generous and helpful comments of Professor John Swales are acknowledged appreciatively.

2. A complete listing of the corpus used is available on request.

3. This selection limitation was not imposed in the Swales study.

REFERENCES


Carrell, P. L. 1983. 'Some Issues in Studying the Role of Schemata, or Background


Fleiss, J. L. 1971. 'Measuring nominal scale agreement among many raters.' Psychological Bulletin 76/5: 378-82.

Garfield, E. (ed.). 1981. Science Citation Index. Philadelphia: ISI.

Garfield, E. (ed.). 1982. Social Science Citation Index. Philadelphia: ISI.


Hartmann, D. P. 1977. 'Considerations in the choice of interobserver reliability estimates.' Journal of Applied Behavior Analysis 10/1: 103-16.


Pitkin, W., n 1979, 'Discourse blocks.' 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.' Ph.D. dissertation, University of Illinois.


