Analysis of the structure of original research papers: an aid to writing original papers for publication

JOHN SKELTON

SUMMARY

Background. An increasing number of people involved in medicine are under pressure to publish research, but there is little understanding of how to describe structured writing.

Aim. This paper aims to describe the structure of original research papers published in the British Journal of General Practice with a view to providing insight into the nature of such analyses, and particularly to help researchers and trainers to write and teach writing more successfully.

Method. A sample of 50 original papers published in the Journal between January 1989 and March 1993 were examined. The papers were subjected to a form of 'move structure analysis', a technique used in applied linguistics; move structure analysis assigns a tentative function to a piece of text, and identifies words/phrases associated with it. To be recognized, moves thus identified had to occur in the same section of the paper in 65% of the corpus, and/or appear in the same order relative to other moves in 50%.

Results. Fifteen moves were identified, four in the introduction, three in the method, and four each in the results and discussion. These moves functioned, for example in the discussion, to state limitations and defend successes; describe achievements; contextualize procedures and findings; and offer recommendations. Frequency scores ranged from 66% to 100%, and order stability scores from 50% to 80%, with three moves being unordered.

Conclusion. It is possible to derive from this study a template for structuring academic medical writing. This template may be built up from the exemplificatory quotations in the text, to provide assistance to educators and less experienced writers.

Introduction

WRITING for publication is an important skill for an increasing number of people in medicine. For some, publication means prestige and career advancement, for others a means of kindling intellectual interest by opening and sustaining dialogue.

There is a growing literature within the field of applied linguistics on the structure of the scientific research paper,[1-4] much of it discussed or alluded to by Swales,[5] and there is a now substantial body of work on the sociology of the scientific research paper.[6-10] Relatively little of this work, however, is on medical

J Skelton, BA, MA, senior lecturer, Department of General Practice, University of Birmingham.

writing and the fact that it is seldom, if ever, cited in academic medicine is probably because it is unknown within medicine. Of studies on aspects of communication within medicine, there is a vast amount on spoken language, particularly the general practitioner consultation — Stewart and Roter include a comprehensive update.[11] However, where attempts at identifying a structural taxonomy are made these tend to rely on observer's impressions rather than analysis. There is less on written language, and almost nothing that touches on the social context of scientific writing, while even Huth's excellent guide has little to say on linguistic structure.[12] Most journals, including the British Journal of General Practice, offer some sort of advice to authors, sometimes (as with this Journal) further referring the reader to the work of the International Committee of Medical Journal Editors.[13] Such advice, however, tells prospective writers what to do rather than how to achieve it. Moreover, the majority of vade mecums (guides) that set out to help with scientific writing are weak, mistaken in matters of language, or exhortatory rather than educational — in effect, full of suggestions that the writer should 'be clear', with misleading indications, or none at all, as to how clarity is to be achieved.

It was assumed that an author seeking publication wishes to stake a claim in the scientific community by shaping knowledge in an appropriate manner. This paper therefore has three aims. First, to offer a description of the structure of papers published in the British Journal of General Practice. Secondly, to offer insight into how the analysis of a text might be approached. Thirdly, and most importantly, to outline a basic template for the structuring of academic work, of value to readers who perceive themselves as having difficulty in writing for publication and which can provide the bare minimum which must be included in an original paper.

Method

The corpus consisted of 50 research articles published in the original papers section of the Journal between January 1989 and March 1993. Only original papers were considered since the constraints on structural variety for this section are stronger: the same is true for most scientific journals. Articles for inclusion in this section are not solicited, and are therefore in competition with other papers and subject to the value judgements of the assessors. For the same reason — the stronger constraints imposed — only papers which contained the traditional sections of introduction, method, results and discussion were included. The great majority of original papers conformed to this structure, it being editorial policy. All 16 original papers in the January to March 1993 issues were included to bias the material studied towards recent practice. All original papers published between 1989 and 1992 were then assigned a number and 34 were drawn at random to make up the balance of the sample.

Move structure analysis

The corpus was then subjected to a form of move structure analysis. This is a technique developed in applied linguistics to analyse spoken or written language. 'Move' as a technical term dates back to Sinclair and Coulthard[14] (a full definition of a

move is a complex matter not attempted in this paper). Move structure analysis is a technique particularly used in the teaching of professional and academic writing, and is typically used as a means of describing what may always be done rather than what must be or is always done. It is important therefore to understand the status of work in this field, including the present paper. It offers an informed description of probabilities, not a list of necessary features.

Move structure analysis tentatively assigns a function to a stretch of written or spoken text, identifies that function with one, or a set of, exponents which signal its presence, and seeks to establish whether or not the pattern identified is a general one, by reference to ostensibly similar texts. If the pattern can be generalized, its status as a move is confirmed. To take a trivial example, one may tentatively identify similarities in what people say on the telephone, which one might offer the functional label 'greeting' and associate with exponents such as 'Hello', 'Hello, can I help you?', the speaker's name/number, and so on. On inspection of a large number of telephone calls it will be observed that this is indeed a consistent pattern which will cover most (but not all) actual telephone greetings, and the existence of the move is confirmed. It will also be observed, as is often the case with moves, that greetings are associated with particular places in the spoken or written text — here, the beginning. However, the conventions for telephone greetings could be articulated by anyone, and to state them is therefore vacuous. It is the strength of move structure analysis that it can identify conventions which are not part of the conscious knowledge of the user, and are not therefore recognized or articulated.

Analysis

A move is in principle identifiable by the common association of function and exponent. For teaching purposes, however, it is important to indicate further not merely whether and how often a move occurs, but where it occurs. In this study, a move was identified as a major feature if the following criteria were satisfied. First, there must have been a pattern of association between function and exponent. Secondly, the feature must have occurred in one section (introduction, method, results or discussion) in 65% of the material studied. These two judgements together constituted the basic measure of frequency. Thirdly, unless the feature had a frequency of at least 85%, it had to be ordered. An ordered move was one which was flanked by the same move before and after (after only in the case of the first move, before only in the case of the last move) on 50% of occasions. This gives a potential measure of order stability.

If the writer is to make his/her own judgements about the strength of constraint which operates with respect to a particular move, information about just how typical it is must be readily available. For these reasons, statements of straightforward probability were considered most appropriate.

Results

As only original papers with an introduction, method, results and discussion structure were eligible for inclusion in the study, six of the original sample were excluded and replaced at random.

A total of 15 moves were recognized, four in the introduction, three in the method, four in the results and four in the discussion (Table 1). These moves were numbered 1–15, chronologically in the order in which they most often appeared, except that moves nine, 10 and 11 were typically interwoven, in no particular order, in the results section. The table does not claim that the moves only occurred in these positions if they occurred at all. Moves five, six and seven were sometimes accompanied by a move, which was here called a tied move (it could not exist on its own,

Table 1. Frequency and order stability of move structure of the 15 moves identified in the original papers.

Move number	Frequency (%)	Order stability (%)
Introduction		
1	76	62
2	100	50
3	70	56
4	100	64
Method		
5	100	74
5b	58	–
6	100	80
6b	32	–
7	86	80
7b	6	–
Results		
8	84	54
9	100	Not ordered
10	100	Not ordered
11	100	Not ordered
Discussion		
12	72	68
13	66	60
14	68	54
15	90	80

invariably appearing with the principal move). The function of the tied moves was to justify the choices of procedure indicated.

Introduction

The function of the introduction was to create a research space, by stating the relevance of the study, contextualizing it in the literature, claiming its novelty, and stating its purpose. The four moves recognized in the introduction were labelled according to Swales,[15] though the criteria for inclusion employed in this study were less elaborate, in order to make the study more approachable. The result was a more conservative set of claims.

Move one asserted the importance of the field of study. Typical exponents were 'important', 'central', 'common', 'expanding', 'evolving', 'growing'. Thus:

'Working paper 4 of *Working for patients* and the subsequent *Improving prescribing* both highlighted the need to refer to the age of a practice's patient population when considering a practice's prescribing patterns. The importance of age as a factor affecting prescribing patterns was subsequently endorsed in a paper by Foster and Frost.'

(1993; p.102)

'Hypothyroidism is a common disorder, a large screening study in the north of England revealing a prevalence of up to 19 cases per 1000 women.'

(1993; p.107)

This move occurred in 76% of the papers studied (that is, a 76% frequency), and in 62% of cases it occurred first (that is, order stability of 62%). An interesting minor variation was, as above, to draw attention to a document in the public domain:

'The government's recent white paper[1] confirms that the primary care team has an important role in lifestyle counselling...'

(1989; p.196)

Where this variation included an exponent such as 'important' (as above) it was recognized as move one; where it did not, it was not so recognized, though it was likely that the mere act of reference to a document as weighty as a white paper was designed to assert importance. The second move was a discussion of previous literature whose function was to contextualize the study in a recognized research tradition. It was invariably present. Its exponents, straightforwardly, were researchers' names and/or superscript numbers in the text. This move could appear several times as various aspects of previous research were picked up and discussed in turn. A minor variation here was to assert significance by reference to another writer, thus combining the first and second move in the same sentence — indeed the presence of the superscript in the previous quotation indicated that this was happening here. In this study, such an occurrence was consistently recognized as a move one followed by a move two. Move three drew attention to a gap in the literature, and asserted that a particular research question required an answer. It was typically associated with an exponent such as 'but', 'however', 'nevertheless', or such as 'little is known'. For example:

'There have, however, been few studies that look at the effects of such a change in the National Health Service.'
(1993; p.52)

'Little is known about psychological mechanisms linking disorder with presentation...'
(1993; p.6)

Move four stated the writer's aim — typically, to fill the gap indicated at move three. Possible exponents included 'aim', 'intention' and 'purpose'. Thus:

'The aim of this study was to investigate the attitudes of general practitioners, health visitors, community midwives and district nurses to the psychological aspects of miscarriage.'
(1992; p.97)

Sixty per cent of introductions in the sample had all four moves, and 50% of the sample had them in the order one, two, three, four. There were occasions, in this and other sections, where there were strong reasons to identify a stretch of text as functioning to fulfil a particular move — especially move four — but without a formally identified exponent; these were not recognized.

Method

The function of the method section was to describe procedures used and assert their credibility.

The three central moves of the method section, moves five, six and seven, were extremely likely to occur, and maintained a strong order. Move five identified the population to be studied, its exponents being, quite straightforwardly, numbers and dates, and labels of occupation, social role or class such as 'doctor', 'patient', 'social' class 1'. This move was always present, and its presence easy to recognize. It could be accompanied by a tied move, which if it occurred, occurred immediately after, or interwoven into, the same point of the text. The function of this tied move was to assert the inclusiveness of the population under study; exponents were 'every', 'all', 'each':

'All successive patients with postherpetic neuralgia of more than three months' duration... were considered for inclusion in this study.'
(1992; p.244)

Move six was always present. It described the procedures used, its exponents were labels for procedures, such as 'questionnaire', and it was associated with a relatively rare tied move which overtly justified choice of procedure by reference to the literature, a pilot study, or consultation procedure:

'After consultation with both general practitioners and specialists, 13 questions were devised.'
(1993; p.111)

'A pilot study was conducted during a four week period two months before the main study.'
(1993; p.61)

Move seven named statistical tests to be used. The exponent was often a subsection labelled 'analysis' or 'statistics', or the name of a statistical test. A tied move justifying statistical choices was rare (frequency 6%). Its only exponents in the corpus were 'as' and 'because'.

Of the corpus 86% had major moves five, six and seven and in 74% of the corpus they appeared in this order. In none of the papers was all three tied moves represented.

The method and results sections were more heterogeneous than the introduction or discussion, because they were strongly content-driven.

Results

This section functioned to describe (but not interpret) data in an ostensibly objective manner.

The four moves of the results section were all extremely likely to occur, and moves nine, 10 and 11 had to occur. Move eight was concerned with adjustments and exclusions from the original population normally given at move five, and exponents — of which there were many in the corpus — included 'withdrew', 'refused/declined', 'were excluded', or more broadly, as in the first example below, a comparison of two figures:

'Of the 34 patients identified 32 participated in the study...'
(1993; p.53)

'A total of 404 were excluded from the study...'
(1993; p.66)

This move occurred with great frequency, and was fairly stable. It always occurred where the cooperation of individual subjects was required for a longitudinal study, for obvious reasons. In a few additional cases (8%) adjustments were included immediately after move five, in the method section.

Moves nine, 10 and 11 were not ordered: move eight was counted as an ordered move in that it occurred after the end of the methods section, and immediately prior to moves 10 and 11. Move nine, the representation of some part of the results in tables, was not part of the main body of the text and therefore had no position in it; the geographical location of the tables on the page was not an authorial decision in any case. Move 10 was a discussion of the data in words, and occurred in every paper together with an assessment of the data (move 11). In effect, these moves happened at the same time. The results section, in fact, was typically either a list of move 10s or a repeated cycle of move 10 plus move 11, until the data in the tables were exhausted — there was evidence of this pattern in every paper in the corpus. For example:

'The use of deputies changed with the time of the call

(Table 3). Daytime calls to all ages of patients were usually covered by the practices themselves, but deputizing visits rose from 11% of calls during the daytime and 10% in the evenings to 26% at night. Telephone advice was relatively stable at 35–42% of all calls throughout the three time periods;'

(1990; p.20)

The pattern exemplified here was typical in three ways. First, the information was already available in the accompanying table, and the main function of the text was not to introduce new information, but to highlight key areas of the table. Secondly, the data were offered in words (move 10) and were interpreted. However, thirdly, there was nothing contentious about the interpretation. Thus, a spread of 35–42% was evaluated as 'relatively stable' while the percentage of daytime calls covered by the practices was sufficiently high (the table made clear it was in fact 80%) to attract the word 'usually'. Strictly speaking, these were subjective judgements rather than scientific conclusions, but they were entirely anodyne.

The use, in general, of the word 'significant' was important in this regard. Where it appeared in the results section it was invariably non-contentious because it meant of statistical (therefore, objective) significance. Assertions of significance, particularly in the discussion section, where it was not statistical significance that was in question, occurred frequently in the earlier papers in the study.

Discussion

The function of the discussion was to contextualize the research, and assert its value.

Move 12 stated limitations and defended successes. Principal exponents were 'caution', 'limitation', 'difficulty', 'excellent', 'successful', 'justification'. The move was common and stable. It took three forms. One possibility was that an overt claim for the success of the study or some part of it was made:

'In this study excellent cooperation was received...'

(1991; p.325)

A limitation to the paper (or more broadly, a sense of anxiety about some aspect of it) may simply have been noted:

'The results of this study should be interpreted with caution... The lack of demonstrable relationships... could be a function of the inadequacy of the test procedures themselves.'

(1993; p.118)

Or thirdly, it may have been presented as an issue encountered and resolved, in which case it was accompanied by an adversative, such as 'however'. The following extract illustrates both of these last two options:

'We initially wondered whether taking blood samples from elderly patients was justified. However, venesection is a relatively non-invasive, safe practice and a procedure that can easily be carried out by a trained practice nurse. We also have access to good laboratory facilities, with caring and cooperative staff. There was however, a logistical difficulty in that the practice is 15 miles away from the local district general hospital.'

(1991; p.497)

Move 13 presented the central achievements of the study, and reflected move four:

'This project has shown that it is possible to set up a system which keeps general practitioners informed of the deaths of their patients.'

(1993; p.72)

Yet, though move four was obligatory, only 66% of the corpus had a move 13 with a clear exponent, as above, such as 'demonstrates', 'shows'. Move 14 contextualized the researcher's procedures and findings in a relevant research tradition, almost always by pointing out that they were in accord, or compared well with those of another study:

'The 86% response rate to the questionnaire obtained in this study compared well with rates in previous surveys... 68% [of practice nurses] had been in post less than one year. This compares with the results of Kerr and colleagues...'

(1993; p.98)

At one level, this move validated the paper by claiming membership of the academic community: at another, it reflected move two, which was also designed to contextualize the study by reference to literature.

Move 15 was common and ordered. It offered recommendations about what should or must happen. These recommendations fell into three broad, overlapping categories which could not be distinguished with certainty. One was a recommendation for some highly specific, potentially achievable course of action, sometimes clinical but often to do with a change to infrastructure or training. A second category was a broad call for more or less specified further research. A third was for unfalsifiable goals to be attained (it would be difficult to determine when or whether the goal had been reached). For example:

'We must also address the possibility, previously largely ignored, that for some people at least, screening can do more harm than good.'

(1989; p.195)

At times, particularly where the recommendation was present only in the final sentence of the paper, it appeared not to represent a serious attempt to change the real world. The conclusion aimed for rhetorical rather than logical persuasiveness:

'A figure of more than three quarters of a million consultations a year involving homoeopathic treatment represents a level of clinical activity that needs to be taken seriously and examined carefully.'

(1989; p.505)

Of papers in the corpus 36% contained all four moves, a low figure which reflected the relative freedom from constraint in the discussion section. However 80%, had at least three of the moves represented.

Discussion

An important limitation to this study is that moves recognized in the results section were at such a level of generality that they are of limited value. This section is driven strongly by content rather than structural convention, as is the actual description of the method in the method section. Given that move structure analysis has nothing to say about content, it may be that a linguistic analysis ought not to attempt anything further in this area. More

importantly, identifying only the commonplace, as is attempted here, is a mirror of competence, not brilliance; it falsifies the complex and subtle nature of good scientific writing — and for that matter, the clumsy nature of poor scientific writing. In writing there is an overriding convention that conventions may be sensitively broken. If this were not so, both writing and reading would be dull beyond description, and the concept of good writing — one common definition of which is that it consists of the unexpected intelligently deployed — would have no meaning.

One question which cannot be resolved by this or similar studies of the published versions of papers is the extent to which the selection and ordering of moves is undertaken by the editorial team rather than the author. This presumably varies. Preparation of work in template form, however, does not merely reduce the task of sub-editors. It helps the author think and organize work in a form which has stood the test of time, and will at least not hinder the chances of a paper being accepted. This study has demonstrated that it is possible to make generalizations which are unfamiliar and therefore educationally valuable about how to achieve structural competence within the social setting of this *Journal* at least. As a resource for professional writers the template may be followed precisely by the less experienced and, judiciously employed, can act as a guide for the more experienced. Findings are in accord with studies such as are detailed in Swales,[5] and with Huth's brief remarks on structure:[12] with local variations they are likely to be broadly replicable in other refereed, academic, medical journals.

There were two interesting subsidiary findings. First, a characteristic feature of the discussion section in scientific writing is its highly speculative nature, and the corpus was no exception. If markers of uncertainty and tentativeness such as 'may', 'seems' and 'possibly', are considered (these are known as 'hedges'[16]) there was some evidence that the discussion section in a scientific paper was at least as speculative as an academic paper in art criticism.[17] Speculation was a pervasive feature of discussion sections and almost entirely absent from other sections. Secondly, it was remarkable how frequently papers ended with a recommendation which was too imprecise to operationalize, or too grand to be implemented by a decision at much lower than ministerial level. Perhaps researchers are uneasy at the prospect of presenting descriptive rather than prescriptive work, and of risking the subsequent charge that their study lacks point. This raises the question of how the discipline of primary care sees itself and seeks to present itself to others. More broadly, medical researchers, and (particularly) postgraduate trainers need to have a way of articulating the structure of professional and academic writing, such as is offered here.

With the exception of this sentence, the paper conforms precisely to the move structure analysis it describes.

References

1. Crookes G. Towards a validated analysis of scientific text structure. *Applied Linguistics* 1986; **7:** 57-70.
2. Peng J. Organisational features in chemical engineering research articles. *English Language Research Journal* 1987; **1:** 79-116.
3. Swales JA, Najjar H. The writing of research article introductions. *Written Communication* 1987; **4:** 175-192.
4. Hopkins A, Dudley-Evans A. A genre-based investigation of the discussion sections in articles and dissertations. *English for Specific Purposes* 1988; **7:** 113-122.
5. Swales JA. *Genre analysis. English in academic and research settings.* Cambridge University Press, 1990.
6. Knorr-Cetina KD. *The manufacture of knowledge: an essay on the constructivist and contextual nature of science.* Oxford: Pergamon, 1981.
7. Gilbert GM, Mulkay M. *Opening Pandora's box: a sociological analysis of scientific discourse.* Cambridge University Press, 1984.
8. Myers G. Texts as knowledge claims: the social construction of two biology articles. *Soc Stud Sci* 1985; **15:** 593-630.
9. Bazerman C. *Shaping written knowledge.* Madison, WI: University of Wisconsin Press, 1989.
10. Myers G. *Writing biology: texts in the social construction of science.* Madison, WI: University of Wisconsin Press, 1990.
11. Stewart M, Roter D (eds). *Communicating with medical patients.* Newbury Park, CA: Sage, 1989.
12. Huth E. *How to write and publish papers in the medical sciences.* Baltimore, MD: Williams and Wilkins, 1990.
13. International Committee of Medical Journal Editors. Uniform requirements for manuscripts submitted to biomedical journals. *BMJ* 1991; **302:** 338-341.
14. Sinclair JM, Coulthard M. *Towards an analysis of discourse.* Oxford University Press, 1975.
15. Swales JA. *Aspects of article introductions.* Birmingham: University of Aston, 1981.
16. Lakoff G. Hedges: a study in meaning criteria and the logic of fuzzy concepts. In: Peranteau P, Levi J, Phares G (eds). *Papers from the eighth regional meeting, Chicago Linguistics Society.* Chicago, IL: CLS, 1972.
17. Skelton J. Comments in academic articles. In: Grunwell P (ed). *Applied linguistics in society.* London: Centre for Information on Language Teaching and Research, 1988.

Address for correspondence

Mr J Skelton, Department of General Practice, The Medical School, University of Birmingham, Edgbaston, Birmingham B15 2TT.