Journal of Applied Linguistics

JAL (PRINT) ISSN 1479–7887
JAL (ONLINE) ISSN 1743–1743

Article

An 'important contribution' or 'tiresome reading'?

A study of evaluation in peer reviews of journal article submissions

Martin Hewings

Abstract

While the process of peer reviewing journal articles submitted for publication has been extensively investigated, particularly in the biomedical field, the language of peer reviews is relatively unexplored. This paper studies evaluation in an electronic corpus of 228 reviews submitted to the journal English for Specific Purposes (ESP). The research focuses on the things (or entities) evaluated and the adjectives associated with these. Entities and adjectives are categorised and quantified in order to ascertain what things are valued by reviewers and the qualities by which they are judged. The findings suggest that reviewers take on multiple roles, at the same time discouraging the publication of work that fails to meet the required standards and offering encouragement to authors and guiding them towards publication. These findings have implications for authors submitting research papers, those who support authors in this process, and journal editors.

KEYWORDS: PEER REVIEW; EVALUATION; PUBLICATION PROCESS; ADJECTIVES; CORPUS

Affiliation

Martin Hewings, English Department (EISU), School of Humanities, University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK.

email: m.j.hewings@bham.ac.uk

JAL VOL 1.3 2004: 247–274

equinox
www.equinoxpub.com

1 The role of peer review in the publication process

Peer review forms part of the decision-making process in a number of scholarly activities, including conference paper submissions, applications for research grants and book proposals. However, it is in the selection for publication of research articles submitted to journals that peer review has received most attention as the main method by which the quality of potential contributions is assessed.

In most academic disciplines, the research article is the primary written outlet by which the results of investigations are recorded and reported; the process of research article writing and revision has become, in itself, the focus of a growing body of research (e.g. Bazerman, 1988; Berkenkotter & Huckin, 1995; Gilbert & Mulkay, 1982; Latour & Woolgar, 1979; Myers, 1990). Assessment of articles and the research which they embody takes place at various stages in the process of research and writing. During the course of research, experimental techniques, data collection, and drafts of papers may be discussed and commented on by colleagues. Once the paper is submitted for publication, peers judge the quality of the research, the validity of the knowledge claims made on the basis of the findings (Gilbert, 1976), and whether conventional style and format are followed in the presentation. After publication, the most common way of offering credit and recognition to a paper is through citation (Cronin, 1984): papers cited are recognised as having made a contribution to the stock of knowledge held by the academic community. Although only one of the mechanisms of research assessment, peer review is key to the process of making research findings accessible to the wider academic community: without acceptance by peers research articles go unpublished. The peer review is, therefore, an important component of the dialogue, negotiation and building of consensus (Kuhn, 1970) which leads to the construction of knowledge. More mundanely, though no less significantly for those involved, peer review is important in the career paths of those working in academia: publication allows scholars to contribute to their discipline and, in so doing, obtain or maintain a job or gain promotion or tenure.

The importance of peer review in the publication process makes it a worthy area of investigation, and a substantial body of literature exists on the subject. The general aim of these studies is to evaluate its effectiveness in controlling the quality of published information. The vast majority of this work has been undertaken within the biomedical field, a fact not surprising given that ensuring the quality of published information in this area can be a matter of life or death (Rowland, 2002). Among the particular areas of interest have been: assessing the quality of peer review (Jefferson et al., 2002); the ethics of the process (Ernst & Resch, 1994; Nylenna et al., 1994); the qualities of reviewers (Chilton, 1999; Goldbeck-Wood, 1998) and the effectiveness of blind (Godlee

et al., 1998; McNutt et al., 1990), open (van Rooyen et al., 1999; Walsh et al., 2000) and online peer review (Wood & Hurst, 2000).

Relatively little work, however, has been reported on the language of peer reviews. This is perhaps not surprising given that they are written for editors and authors alone, and not intended for wider scrutiny. Questions of confidentiality, therefore, prevent their wider circulation and consequently their availability for research. However, within the growing interest in the analysis of academic genres, attention has been turned to 'occluded genres' (Swales, 1996), such as the peer review, which are generally hidden from public view but are none-the-less important in the research process. Analysis of the language of peer reviews can reveal insights into the relationship between reviewers as gatekeepers of standards and conventions within a particular discipline, and writers who seek to convince that they are able to produce writing whose content and presentation conform to the communication conventions established by the discipline. It contributes, therefore, to our understanding of 'the sanctioned social behaviours, epistemic beliefs, and institutional structures of academic communities' (Hyland, 2000: 2).

Useful in this analysis are the categories of 'centre' (or 'core') and 'periphery'. These notions have been widely used in analyses of relations between those on one side or the other of a centre-periphery divide between institutions, geographical regions, or social groups (Fox, 2003). In the context of English language teaching, Phillipson (1992), Pennycook (1994), and others, have distinguished between institutions and communities in the centre – in more developed nations where English is spoken as a first language – and those in the periphery – in less developed nations where English is spoken as a second or foreign language. Referring to the academic community, the notions have been used in discussions of ways in which members of the community participate (Lave & Wenger, 1991); that is, concerning the degree of engagement that a scholar has with the defining practices of a particular community (Canagarajah, 2002; Casanave & Vandrick, 2003a). The majority of peer reviewers of research articles are representatives of the 'core' or 'centre' in that they have already shown their ability to meet the requirements of the disciplinary community through international publication, although this does not suggest, of course, that core scholars themselves do not have to submit to the peer review process. Scholars may be on the 'periphery' for a number of reasons. They may be novices, new to academic publishing, or unfamiliar with publication procedures; they may be non-native speakers of English – the predominant language of international academic publication (Graddol, 1997) – where high standards of written English are required; they may operate in academic cultures where the patterns of scholarly work, debate and written expression are different from those in the core; or they may lack the kinds of resources, including access to recent literature or computers, that facilitate international publication. Much of the work on the journal publication

process conducted from an applied linguistic perspective aims to describe and evaluate the particular difficulties of periphery scholars in getting published with the eventual aim of easing the path to publication, either by improving research writing training for periphery scholars or to encourage fundamental changes in the publication process itself. For example, Canagarajah (1996; 2002; 2003) explores the experience of scholars from developing countries publishing in international journals in English. Particular groups of non-native English-speaking scholars are investigated by: Burrough-Boenisch (2003) – Dutch scientists; Curry and Lillis (2005) – Hungarian, Slovakian and Spanish psychologists; Gosden (2001; 2002; 2003) – Japanese doctoral students of applied physics, chemistry and cell biology; Flowerdew (1999a; 1999b; 2000; 2001) – Hong Kong academics in various disciplines and Kourilová (1998) – Slovak medical doctors submitting to biomedical journals. In addition, Benfield and Howard (2000) compared comments on language in peer reviews of papers submitted by native and non-native speakers to journals in the field of thoracic and cardiovascular surgery.

One important lesson to be drawn from this growing body of work is that the process of publishing journal articles is complex and variable. For example, the participants in the review process may have multiple aims and motivations. While the ultimate goal of the author may be to publish, in submitting a paper to an international journal they may wish to obtain feedback on their work by experts in the field through peer review (feedback which may be hard to come by in the local working environment of some scholars), and may use this in guiding their research or in revising the paper for eventual publication elsewhere. Reviewers, as well as acting as gatekeepers, may see it as their role to guide authors towards publication, effectively acting as anonymous mentors in the development of an author's publishing career. The research to be reported here suggests that reviewers do indeed take on multiple roles, and this theme will be taken up later in the paper. It should also be noted that the peer review process is different in its detail from discipline to discipline and from journal to journal. Bordage and Caelleigh (2001: 904) note that:

> whereas journals in all branches of science share the core ethos of peer review, it has evolved in diverse ways to best fit the environments and circumstances of the various sciences and disciplines.

It seems likely that, just as the form of peer review varies across disciplines, so too will the language of peer review. While much of the relevant work has so far been conducted on science journals, the present study extends this to an investigation of a journal within the social sciences. It should be emphasised, however, that it focuses only on reviews submitted to one journal in one discipline, and it is not suggested that the findings can be generalised outside this

limited area. However, as an exploratory study it seeks to identify possible areas for further investigation, including in other journals and other disciplines.

2 Previous work on the language of peer review

The most systematic analyses of the language of peer reviews written in English are those conducted by Kourilová (1998) and Gosden (2001; 2002; 2003). Kourilová (1998) analysed 80 peer reviews received by Slovak biomedical researchers writing in English. She focused primarily on: (i) a set of discourse features – critical items, commands, hedges, compliments, impersonality devices, the involved textual dimension (personal verbs, person pronouns and contractions), doubt and suspicion, and assumptions; and (ii) the targets of criticism, grouped into areas of omission, failure of economy (of words, thought, etc.), language, design deficiencies, unjustified conclusions, formal mistakes (typing errors, numbering, etc.), presentation shortcomings, ethical considerations, and statistics. Her main conclusion is that politeness conventions in peer reviews are different from those in most other forms of academic writing, with criticism more 'on-record' and less hedged or mitigated. There were over four times the number of blunt criticisms in her reviews than hedged criticisms. This, she argues, is a consequence of the difference in power between writer and reader for peer reviews and research articles, and the fact that peer reviews are anonymous and not publicly available, even though they may be made available in an unmediated form to authors. In addition, the targets for criticism identified by Kourilová include not only problems of style and grammar, but also more subtle language choices such as the selection of appropriate modality in order to express generalisations and claims with appropriate levels of certainty or caution.

Gosden (2001; 2002; 2003) has studied various features of a corpus of 40 peer reviews written by 15 Japanese and 25 native English-speaking reviewers on papers submitted to a hard science *Letters* journal. In such a journal speed of publication is paramount and paper length is restricted: in the journal examined by Gosden the time from submission to publication was one to two months, and papers were restricted to 2000 words. Some of the papers were designated acceptable with revision, and others unacceptable. Gosden (2001; 2003) classified referees comments according to topic, giving the percentage of the total comments for each sub-category as follows: discussion (33.8 per cent); technical detail (26.9 per cent); claims (19.8 per cent); references (12.5 per cent) and format (7.0 per cent). Additional comments were on difficulties with English language, although the majority of these referred only to general problems. Gosden (2002) looked at the content of sentence-initial Themes in the reviews in order to gain insights into the motivation for referees' comments. He found that

some two-thirds of Themes focused on 'the degree to which referees believed manuscript authors had been successful in making a convincing presentation of their research findings intended to target readers' (2002: 11).

Benfield and Howard (2000) looked at articles about lung cancer submitted to the journal *Annals of Thoracic Surgery* in Volumes 65 to 68 (published between 1998 and 1999), focusing on the 'language burdens of NNS [Non-Native English Speaking] authors'. Of the 50 articles, 27 were by NNS authors and 23 by Native English Speakers (NS). Reviewers' comments on 'communications skill' were identified and categorised according to their focus: Language (e.g. 'Some of the English structure and spelling needs revision', 'The term metastatic lung cancer is ill advised'); Organisation (including comments on Discussion and Results or methods e.g. 'The discussion is not precise, but rambles'); and Writing quality (e.g. 'This section needs to be re-written so that the reader can understand what the authors are trying to explain.'). Most comments were in the Language category, and the most significant differences between the NS and NNS manuscripts were in the categories of Language and Writing quality. In a sample of the manuscripts, language errors were also identified by a language professional. 263 errors were noted, compared with 51 by the peer reviewers who 'focused more often on word choice, and the need for correction of grammatical errors in a general way'. On the basis of their work, Benfield and Howard have called for steps to be taken to improve editors' and reviewers' awareness of the added burden English imposes upon NNS authors and, for the field of thoracic surgery, recommends a mentoring service for NNS authors in which NS thoracic surgeons make themselves available for assistance.

Mention should also be made briefly of the work of Canagarajah and Flowerdew. Canagarajah has written extensively on writing from 'the periphery' and, in passing, he reports on and explores his own publishing experience, including the feedback received from reviewers. In Canagarajah (2002) he reflects on details of the reviews of his submission to the journal *Written Communication*. For example (2002: 22), after giving extracts from three reviews received, he observes:

> In hindsight, I see many of these comments as suggesting the typical objections one can have against a more personal, narrative, contextually grounded mode of research reporting. The first referee was asking me to make my argument more explicit and my presentation more analytical. The more indirect and embedded form of my argumentation also earns the censure that the paper is not tightly organized.

Although not based in text analysis, Flowerdew's work (1999a; 1999b; 2000; 2001) is relevant here because, through interviews and questionnaires, the perceived problems of publishing in English by Hong Kong scholars with

Cantonese as a first language, and the attitudes of journal editors to contributions by non-native English speakers are explored. One of Flowerdew's findings (2001) was that among the journal editors interviewed

> there was a general sympathy toward variation in discourse style and nativized varieties [in NNS contributions], the key criterion of acceptability being whether or not the readership would be likely to be able to understand it. (Flowerdew, 2001: 145)

Beyond this work, little detailed analysis of peer reviews in English has been undertaken from a text analytical perspective. Until a larger research base is built up – and the work reported here aims to contribute to this – we are not in a strong position to consider what insights peer reviews can shed on the accepted communication practices of a discipline, nor to explore disciplinary variation.

3 Research aims

The research reported in this paper is intended to add to our developing understanding of the language of peer reviews as a whole, and in particular in one disciplinary field, applied linguistics, and forms part of an ongoing larger investigation of reviews of papers submitted to the international journal *English for Specific Purposes*. It attempts to identify what reviewers value highly (and, conversely, do not value highly) in their reports, and analyses aspects of the language used to indicate this. The specific objectives are to identify and categorise:

(i) the entities which are evaluated;
(ii) the qualities used to judge these entities.

Parallels can be seen in these objectives and those in work by Hunston (1993) and Thetela (1997) on evaluation in academic research articles. It is assumed that the identification of entities selected for evaluation and values ascribed to them can provide evidence on what the discipline, through reviewers as its representatives, considers as having more or less value. In addition, examination of the form in which this judgement is expressed may provide insights into the relationship between reviewers and authors.

While evaluation can be achieved by a wide variety of linguistic devices (for a fuller discussion see Thompson & Hunston, 2000), a very important and frequent means is the use of *evaluative adjectives*, and these will be the focus of investigation in this paper. Two broad semantic groups of adjectives are often identified: qualitative adjectives, which identify the quality of an entity, and classifying adjectives, which identify the class to which an entity belongs. What are referred to here as evaluative adjectives, which denote judgement of an entity, are a subclass of qualitative adjectives.

Two additional terms will be used in this paper. An instance of an entity evaluated by an evaluative adjective will be referred to as an 'adjectival evaluative act'. The term 'act' is used here (after Austin, 1962 and Searle, 1969) to highlight that what is 'done' is an act of evaluation. The thing that is evaluated by an evaluative adjective in an adjectival evaluative act will be referred to as an 'evaluated entity'. In examples of adjectival evaluative acts, evaluated entities will be underlined and evaluative adjectives italicised. For example:

(1) The writer has a tendency to state truisms, make *unsupported* assertions and *unwarranted* prescriptions. (0622)

(2) The introduction is rather *disjointed* and *vague*… (0492)

The code given at the end of examples refers to the peer review from which they were extracted.

4 Data and methods

4.1 Corpus details

The corpus consists of 228 reviews of papers submitted to the journal *English for Specific Purposes* during the period between December 2001 and April 2004. These were compiled into an electronic corpus of around 160,000 words. All reviews submitted to the UK office of the journal during this period were potentially available for inclusion in the electronic corpus. Some, however, were no longer on file and no attempt was made to recover these. Reviewers for whom email addresses were available were contacted requesting permission to use their reviews anonymously in the research and to add them to the electronic corpus. All reviewers who responded gave permission, while the reviews of those who did not reply were excluded on the basis that lack of reply may indicate unwillingness to give permission. Some of the reviews were already available in electronic form, while paper copies were converted to electronic form by scanning or retyping. A number of second reviews gave a minimal recommendation to publish with no further comment, and these were excluded from consideration. All parts of each review were included with the exception (where these existed) of: text at the beginning of the review which gave information such as the title of the paper, the name of the author, or the recommendation according to the list of options provided by the editor (see below); and references included at the end of the review. The corpus includes both 'First reviews' and 'Subsequent reviews'. First reviews are those written by reviewers who had not seen the paper before. In most cases these reviews are produced on first submission of the paper. However, if a paper is revised and resubmitted and this is sent to a new reviewer, these are also counted as first reviews. Subsequent reviews are of

resubmitted papers, and written by reviewers who had read earlier versions of the same paper. Details of the number of first and subsequent reviews and their length are given in Table 1. The great majority of reviews in the corpus were first reviews, and on average these were somewhat longer than subsequent reviews. Perhaps not surprisingly, particularly long reviews were mainly first reviews, while particularly short ones were subsequent reviews.

	First reviews	Subsequent reviews	All reviews
Number	194	34	228
Number of words	136, 945	21,372	158,317
Average length (words)	706	629	694
Longest (words)	4772	1925	4772
Shortest (words)	115	73	73

Table 1: Peer reviews in corpus: numbers and length

The 228 reviews were written by a total of 56 reviewers. Most reviewers wrote more than one review, and some substantially more than that. The highest number of reviews written by one reviewer was 12.

Reviewers are provided with two documents by the journal editor. One lists a set of recommendations which reviewers are asked to select from ranging from (1) 'Publish as an article in its present form or with minor stylistic changes', to (7) 'Not recommended for publication'. Some reviewers indicated a numbered recommendation on their reports, although many gave a more general recommendation without specifically referring to a number on the list provided. For the purposes of this research, the reviews were grouped into one of three categories: those recommending that the paper should be published in its current form or with very minor changes ('Publish' in Table 2); those recommending that the author(s) should revise and resubmit the paper ('Revise and resubmit'); and those recommending that the paper should be rejected ('Reject'). The majority of reviews in the corpus recommended 'Revise and resubmit'. Details are given in Table 2. Of particular note is that reviews recommending 'Revise and resubmit' are substantially longer than those either rejecting or accepting papers, a point which will be returned to below.

	First reviews	Subsequent reviews	All reviews
Publish	25 (432)	14 (285)	39 (378)
Revise and resubmit	131 (835)	17 (924)	148 (845)
Reject	38 (479)	3 (694)	41 (490)

Table 2: Number of peer reviews in corpus according to recommendation by reviewers (average length in words)

The second document provides a set of guidelines for evaluating the paper, although authors are encouraged to use further criteria for evaluation if they wish to. As these are potentially of particular significance in this research, they are listed in full below. Experience suggests that these are fairly typical of the guidelines sent out by journals in the field of applied linguistics at least.

1 Is the topic of potential interest to professionals active in the field of ESP? (And please remember here that *ESPj* serves a wide range of international readership.)
2 Does the article represent an original contribution of knowledge, a perceptive restructuring of existing knowledge, or the discussion of an idea with information and references on how to learn more about the topic?
3 Does the author demonstrate a high level of familiarity with and sophistication in the subject matter field? Are references appropriate and sufficiently extensive?
4 If experimental research is reported, do the research plan, methodology, and statistical treatment of results meet accepted standards? Are the verbal conclusions which are drawn accurate and justified by the data?
5 Is the author too succinct in his/her presentation? Too wordy? Is the overall length of the article appropriate to the material treated? Is the material well organised and the writing style interesting?

In only a handful of cases did reviewers structure their reviews as a list of five separate sections dealing with each of these criteria in turn, although most reviewers made passing reference to some of these criteria. On the same document reviewers are asked to write comments to the editor but in such a way that they could be forwarded to the author 'without extensive censorship by the editors', and in a tone that is 'frank and firm but collegial!'. The journal operates a policy of double-blind review: that is, reviewers are not given the name of authors, and authors are not given the names of reviewers. For their information, a reviewer of a particular paper is sent anonymous copies of other reviews of the same paper. This, however, is a relatively recent practice, and it was not clear how many of the reviews in the corpus were written after it was instituted.

While it is hard to characterise with precision the ethos of the journal, and how this might be different from that of other journals, a number of relevant observations can be made. First, a large majority of reviewers in this study were existing members of the editorial board of *English for Specific Purposes*, or have since gone on to join the board, and all had themselves published academic work, including research articles, many in the pages of *English for Specific Purposes* itself. Second, the community of those engaged in ESP is relatively small in comparison to many other disciplines. Authors and reviewers may well know of each other's work, and may know each other personally (although,

of course, their identities are not known during the review process). Third, many reviewers are engaged not only in research but also are (or have been) engaged in teaching academic writing, particular in higher education institutions. Factors such as these may have an impact on the language of the peer reviews analysed, and will be discussed further below.

4.2 Methods

Once the electronic corpus was compiled, a number of steps were followed in order to produce a classification of evaluated entities and evaluative adjectives. First, *Wordsmith Tools* (Scott, 1996) was used to produce a word frequency list. Out of this, all evaluative adjectives were identified manually. Concordance lines were then generated for each of these adjectives, and the evaluated entity identified for each occurrence. At this stage, a number of adjectival evaluative acts were excluded from further consideration. These were of two main types. The first were instances where the evaluated entity was not in fact an aspect of the paper, the research reported therein, or the author. For example, the adjective *succinct* in

(3) For a succinct discussion of different uses of it-clauses see Martin, et al., 1997. (0671)

evaluates previous work rather than the submitted paper. The second were instances where no entity was identifiable to which could be attached a simple and generalised label. These occurred in two main grammatical patterns: *it + be + adjective + clause* (see Example 4) and *I + link verb + adjective* (see Example 5). In

(4) …it would be very *interesting* (and perhaps *important*) to know whether the author's analysis of the data supports Mauranen or not. (0984)

interesting evaluates 'knowing whether the author's analysis of the data supports Mauranen or not' (and implicitly criticises the author for not making this observation), and in

(5) I find it initially *confusing* that only nine readers are referred to… (0752)

confusing evaluates 'the fact that only nine readers are referred to'. Once these were omitted, there remained 1121 adjectival evaluative acts, which included 213 different evaluative adjectives.

In some cases, different lexical items appeared to be referring to the same kind of evaluated entity; for example, *paper*, *article*, *study*, *work*, *piece*, *submission*, *manuscript* and *text*; and *references*, *reference list*, *list of references*, *bibliography*,

and *sources*. In counting evaluated entities, these were amalgamated into, respectively, the *paper* and *references* entities.

Two analyses were then conducted, the findings of which are presented in the next section.

(i) Each type of evaluated entity (e.g. *paper*, *references*) was counted. To make the analytical process more manageable, where there were fewer than three occurrences, these entities were excluded from further consideration. The remaining entities were then categorised into a number of *entity classes* according to their similarity of meaning and, for ease of reference, one of the entities selected as an entity class label (PAPER, EXPRESSION, etc.). In addition, a distinction was made between those entities to which was attached a positive evaluation (+), and those to which was attached a negative evaluation (-) by reviewers. In the vast majority of cases this judgement was straightforward. For example,

(6) The research reported here is genuinely *original*… (1282)

(7) The closing paragraph is *excellent*… (0382)

were judged to be expressing positive evaluation, while

(8) The writing style is also *awkward* in places… (0711)

(9) The design of the study is rather *weak*… (0401)

were judged to be expressing negative evaluation. There were, however, some more problematic cases. A very common pattern observed in the corpus is for something positive to be said about an entity followed by a criticism. This is exemplified in:

(10) The descriptions of what happened in these two courses are *intriguing* but at present quite *general* and *theoretical*. (0691)

While the overall effect of this sentence seems to be to provide a negative evaluation of the entity 'descriptions', I have classified *intriguing* as providing a positive evaluation, and *general* and *theoretical* as providing a negative evaluation. It is also worth noting that a particular adjective may be used to provide both positive and negative evaluations. For example, the adjective *succinct* would normally be thought of as indicating approval, and does so in:

(11) It [the writing] is suitably *succinct*… (0783)

However, in

(12) I believe the literature review could be made a little more *succinct*. (0711)

it is part of a negative evaluation of the literature review in that the reviewer is indicating that it *lacks* succinctness.

It should be noted that no attempt is made here to place each adjectival evaluative act within its wider context. For example, a description of a section of the paper as 'irrelevant' may be, in one review, a reason given for rejecting the paper, while in another a part of a constructive suggestion on how the paper might be improved, but no account is taken of this here. This does not, however, detract from the purpose of this part of the analysis: to identify those entities which reviewers considered worthy of evaluation, and which were evaluated positively or negatively.

(ii) The evaluative adjectives were grouped according to the broad type of quality they represent. This was done partly on the basis of intuition, but also with extensive reference to thesauruses and dictionaries in order to identify adjectives with similar meanings. A single label (INTEREST, SUITABILITY, etc.) was then given to each group for ease of reference. Each label is intended to represent the core of a range of related qualities expressed by the adjectives placed within that group. The positive (+) adjectives in each group indicate the presence of these qualities, while the negative (-) adjectives indicate their lack. Most adjectives can be judged to be inherently positive (e.g. *interesting*, *good*, *clear*) or negative (e.g. *tedious*, *odd*, *confusing*) out of context. Others, however, (e.g. *anecdotal*, *implicit*) can only be interpreted as positive or negative in context, and in these cases the corpus was investigated to ascertain whether they were used in a predominantly positive or negative way, and then classified accordingly. The frequency of occurrence of all evaluative adjectives was recorded.

No attempt has been made in this study systematically to identify recurring organisational structures in the peer reviews, such as might be undertaken with a 'move-step' analysis of the type pioneered by Swales (1981). However, being able to locate adjectival evaluative acts within functional components of the reviews is of potential interest, and will be the subject of further research.

5 Findings

5.1 Analysis of evaluated entities

The classification of evaluated entities into nine entity classes is given in Table 3. Clearly, this classification cannot be taken as definitive: different judgement could be made of some of the placements of entities into classes and, indeed, the entity classes themselves. The information is, therefore, offered as one possible interpretation of the data. The figures in brackets in the centre column show first the number of occurrences of each entity and then the number of positive and the number of negative evaluations of this entity. So, for example, the entity *paper* (and its synonyms *article*, *manuscript*, etc.) was evaluated using an adjective 177 times, 124 of these positively and 43 negatively. The right-hand

column gives figures for the entity class as a whole; first the percentage of the total adjectival evaluative acts, and then the proportion within the class that were positive and negative. So, for example, 22 per cent of all the adjectival evaluative acts were within the entity class PAPER, and 77 per cent of these were positive evaluations.

Entity class	Entities (total no. of occurrences: +/-)	Percentage of total (proportions +:-)
PAPER	paper (article, manuscript, etc.) (177: 124/ 43), contribution (61: 52/9), research quality (11: 7/4), length (7: 7/0), level (5: 5/0), content (4: 2/2), attempt (3: 2/1) (Total – 258: 199/59)	22% (77:23)
EXPRESSION	'specific wordings' (102: 12/90), section (part) (34: 14/20), expression (wording, phrasing, choice of words, etc.) (23: 0/ 23), presentation (19: 10/9), style (writing) (18: 10/8), sentence (11: 0/11), paragraph (8: 1/7), punctuation (6: 0/6), 'use of language' (pronouns, metadiscourse, verb tenses) (6: 0/6), title (5: 0/5), editing (5: 0/5), language (English) (5: 0/5), abstract (3: 0/3) (Total – 242: 49/ 193)	20% (20:80)
CLAIM	claim (generalisation, statement, assertion) (61: 12/49), discussion (59: 24/35), point (28: 20/ 8), argument (case, position, view) (17: 5/12), suggestion (recommendation) (14: 6/8), comment (commentary, observation) (13: 5/8), conclusion (implication) (13: 6/7), explanation (interpretation, speculation) (10: 1/9), assumption (5: 0/5), reason (justification) (5: 2/3) (Total – 225: 81/144)	19% (36: 64)
ANALYSIS	finding (result) (44: 25/19), account (report, description) (33: 16/17), analysis (30: 17/13), term (18: 0/18), concept (idea, consideration) (17: 11/6), distinction (classification, categorisation) (13: 4/9), relationship (connection) (10: 2/8), table (9: 1/8), treatment (8: 4/4), definition (7: 0/7), figure (diagram, chart) (6: 3/3), comparison (5: 5/0), insight (5: 5/0), list (5: 3/2), appendix (4: 3/1), details (4: 0/4) (Total – 218: 99/ 119)	18% (45: 55)
GOAL	research question (21: 14/ 7), topic (subject) (21: 20/1), approach (13: 5/8), focus (emphasis, perspective) (9: 5/4), aim (goal, purpose) (9: 3/6), issue (8: 6/2) (Total – 81: 53/ 28)	7% (65: 35)
EVIDENCE	data (17: 6/11), example (exemplification) (14: 7/7), sample (10: 0/10), evidence (10: 3/7), corpus (7: 1/6), information (6: 3/3), material (5: 2/3) (Total – 69: 22/ 47)	6% (32: 68)
LITERATURE	bibliography (references, list of references, etc.) (29: 12/ 17), literature review (literature survey, overview of the literature, etc.) (23: 16/7) (Total – 52: 28/24)	4% (54:46)
PROCEDURE	procedure (method, methodology) (14: 5/9), rationale (5: 3/2), theory (4: 3/1) (Total – 23: 11/12)	2% (48: 52)
KNOWLEDGE	knowledge (awareness, command, familiarity, grasp, understanding) (17: 10/7) (Total – 17: 10/7)	1% (59: 41)
Totals	1185: 552/ 633	100 (47: 53)

Table 3: Classification of evaluated entities

The PAPER category includes the paper itself, and entities related to the characteristics of the work as a whole: its contribution to the field, its originality, the overall quality of the research reported in it, etc. Over three-quarters of the adjectival evaluations of PAPER entities were positive. The vast majority of evaluations in this category were of the *paper* (and its synonyms) and of the *contribution* it makes, and over three quarters of these were positive evaluations. The most common evaluative adjectives for *paper* were *interesting* and *relevant*, and for *contribution* were *original* and *interesting*, as in:

(13) It is a very *interesting* paper for a number of reasons… (0042)

(14) I think the paper can make an *important* contribution to the field of ESP. (0701)

Typically, evaluations of *paper* and *contribution* were used early in reviews, forming part of a broad positive assessment which then might be expanded on, although which does not preclude subsequent criticism, or indeed rejection. Two representative review openings follow:

(15) This is an *interesting* paper, but I cannot recommend it for publication in ESPJ for a number of reasons. The topic of the paper is appropriate for ESPJ – academic listening – and the research is original. However, the writer does not demonstrate sufficient familiarity with the literature on listening… . (0151)

(16) This paper deals with a very interesting subject, one that is likely to be relevant to readers of the ESPJ. It also represents an *original* contribution to the field. However, it does have a number of weaknesses. The organization of the article is not sufficiently clear. No distinction is made… . (0894)

Within the EXPRESSION category three sub-groups of entities can be identified. The largest group numerically is evaluation of what are labelled here 'specific wordings'. Typically, a section of the paper was copied into the review in quotation marks, on which the reviewer provided an evaluative comment. For example:

(17) Page 6. Line 11. The 'more or less' is *odd*, or in the wrong place. (0093)

(18) The comment 'only the abstracts in the soft knowledge disciplines...' is both rather *sweeping* and *inappropriate* when the evaluation of the work of others is discussed more fully on p12. (1341)

Only rarely did the reviewers attach a positive evaluation to a specific wording; 90 of the 102 evaluations were negative. The majority of evaluations of specific wordings related to the accuracy of the language used, although a small number questioned the validity of statements made in a quoted section of text. No distinction has been made between these, although it could be argued that negative

evaluations of validity would be more appropriately placed in the CLAIM entity class. It is worth bearing in mind, then, that the figure for negatively evaluated CLAIM entities given below might under-represent their actual number. The second group is evaluations of particular parts of the paper: *paragraph*, *section*, *sentence*, *title*. Again, the great majority of these were negative: 43 of the 58 in total. However, 14 of the 15 positive evaluations were of *section*, and with only one exception, *paragraphs*, *sentences*, and *titles* were all evaluated negatively. In the third category are the remaining, language-related entities: *expression*, *presentation*, *punctuation*, *style*, etc. *Editing*, *expression*, *punctuation*, and '*use of language*' (a category including evaluations of use of pronouns, metadiscourse, and verb tenses) were all evaluated negatively, 40 instances in total, as in:

(19) More instances of *awkward* expression, etc, are noted on the hard copy… (1111)

(20) The main improvement needed by this paper is cleaning up the decidedly *erratic* punctuation. (1591)

Language, *presentation* and *style* have almost equal numbers of positive and negative evaluations. For example:

(21) I personally like the direct, *concise* style of writing, which I find easy to read. (0411)

(22) The information is presented in a somewhat *unorganized* and *repetitive* style, … (0851)

In the great majority of cases, then, particular parts of the paper were focused on in order to provide negative evaluations. Similarly, aspects of language and presentation were highlighted in order to criticise them.

In the CLAIM category are entities related to the process by which authors go beyond reports of the literature, findings, etc. and offer *claims*, *comments*, *conclusions*, *explanations*, etc. as part of the argument presented in the paper. The majority (64 per cent) of the adjectival evaluations of these entities were negative, and of particular note is the high proportion of negative evaluations of *assumption*, *claim* and *explanation*. Claims, for example, were said to be *anecdotal*, *impressionistic*, *thin*, *unacceptable*, *unsubstantiated* and *weak*. Where positive evaluation was given, it frequently preceded reservation:

(23) Many of these claims are *interesting* and may well be *true*, but are a bit difficult to sustain on the basis of [the available data]. (0401)

The entity *discussion* has more equal proportions of positive and negative evaluation, and the entity *point* was mainly (in 20 out of 28 cases) evaluated positively as, for example, *interesting* or *important*:

(24) A *new* and very *interesting* point is introduced in the second to last sentence… (0704)

The ANALYSIS category includes entities which are related to the analytical process (e.g. *analysis*, *concept*), background to this process (e.g. *account*, *definition*), the statement of findings (e.g. *finding*, *insight*), and the form in which analysis or findings are presented (e.g. *figure*, *table*). No clear pattern of positive or negative evaluation can be identified here, with most entities receiving both. It is worth noting, however, that *term* (in most cases it is the use of a term that was evaluated rather than the term itself) and *definition* were together evaluated 25 times, all of them negatively. These might be highlighted as particular areas of apparent difficulty for authors.

In the four categories discussed so far are 80 per cent of the total evaluated entities. The remaining five categories are much smaller and will, therefore, be only briefly commented on. In the GOAL category are entities related to the general subject matter or purposes of the research reported. The majority of evaluations in this category were positive, with the two most frequently evaluated entities, *research question* and *topic* together evaluated 42 times, 34 positively. In the EVIDENCE category are entities related to the data analysed or the evidential support provided for claims, with the majority of adjectival evaluations being negative. Of the 17 evaluations of *sample* and *corpus*, for example, only one was positive, with most critical comments concerned with the *small* size of the sample, etc. used. In the LITERATURE category are entities related to the use of literature in the main text or the presentation of bibliographical information. While no clear pattern emerges, there was a tendency to offer positive evaluations of the extent of the literature referred to, but negative evaluations of the way in which bibliographical information is presented. In the KNOWLEDGE category are entities related to the reviewer's assessment of the intellectual qualities of the author. An author might, for example, be said to show a 'good knowledge of the literature around the subject' (0933). A small majority of adjectival evaluations of entities in this category were positive. Finally, the PROCEDURE category includes entities related to the methods of the research or the theory on which it is based. There were almost equal numbers of positive and negative evaluations of entities in this category.

5.2 Analysis of evaluative adjectives

The categorisation of evaluative adjectives into eight groups, and their division into + and – adjectives is presented in Table 4. The number of occurrences of each adjective is given in brackets in the left- and right-hand columns. To make the presentation clearer, adjectives with only one occurrence are omitted from

the table, although figures in the table include these, and some are referred to in the report of findings below. In the centre column figures are given for the percentage of the total number of evaluative adjectives within each group (e.g. 25 per cent were INTEREST adjectives), and the proportion of positive (+) to negative (-) adjectives (e.g. 95 per cent of the adjectives used in evaluative acts in the INTEREST group were + adjectives). Adjectives in bold type were also used in the guidelines provided to reviewers (see above). For clarification, it should be noted that placing an adjective in the + or – category does not mean that evaluative acts using this adjective were necessarily all positive or all negative. For example, while 'interesting' was judged to be a quality that reviewers approved of and some entities were evaluated positively using this adjective, other entities were negatively evaluated using it because they were considered to lack this quality.

Adjectives in the INTEREST group are to do with qualities such as interest, innovation, enjoyment and informativeness. These adjectives form the largest group (25 per cent of the total), and this group also includes the first and fourth most frequent evaluative adjectives in the corpus, *interesting* and *original*. The frequency of occurrence of these two adjectives is unsurprising given that 'interest' and 'original' are mentioned in the first two points on the guidelines sent to reviewers. The adjectives selected suggest that the reviewers valued entities that offer something new and informative (e.g. *innovative*, *novel*, *original*, *refreshing*, *unexpected*, *unusual*, and conversely *conservative*, *old*, *predictable*, *unsurprising*) and that is able to stimulate the reader by encouraging them to consider things in different ways (e.g. *enlightening*, *fascinating*, *interesting*, *intriguing*, *stimulating*, *thought-provoking*, and conversely *uninteresting*). While academic papers are not read by most with the aim of being entertained, the reviewers also seemed to value entities which bring enjoyment (e.g. *appealing*, *attractive*, *enjoyable*, and conversely *tedious*, *tiresome*).

Many SUITABILITY adjectives (17 per cent of the total) are to do with the reviewers' judgement of entities as having high or low quality measured against some standard perceived by the reviewer and, therefore, whether they are acceptable or not to the discipline. Some are general evaluators of quality (e.g. *good*, *excellent*, *strong*(*est*) and conversely *bad*, *poor*, *weak*(*est*)), while others refer to acceptability or appropriacy (e.g. (*in*)*appropriate*, (*un*)*suitable*, (*un*)*satisfactory*) and relevance ((*ir*)*relevant*). Note that the reviewers' guidelines specifically refer to appropriacy in connection with references and the length of the article. Unexpected in this group was the high frequency of *odd*. Entities most frequently judged to be *odd* were features of language. Reference was made either specifically to a section of the text being odd, with the section quoted in the review, for example:

+ adjectives	Group	- adjectives
interesting (199), **original** (44), innovative (7), fascinating (6), intriguing (5), unusual (5), stimulating (4), ambitious (3), attractive (3), illuminating (3), rare (3), novel (2), pertinent (2), thought-provoking (2), unexpected (2), [and 7 others with 1 occurrence] (Total = 297)	1 INTEREST (25%) 95:5	tedious (3), uninteresting (3), unsurprising (3), conservative (2), old (2), [and 2 others with 1 occurrence] (Total = 15)
good (39), **appropriate** (34), relevant (23), suitable (15) excellent (14), effective (3), reasonable (3), satisfactory (2), strongest (2), successful (2), [and 1 other with 1 occurrence] (Total = 138)	2 SUITABILITY (17%) 64:36	odd (21), weak (15), stronger (9), **inappropriate** (6), unfortunate (6), poor (5), unsuitable (4), irrelevant (3), inadequate (2), [and 8 others with 1 occurrence] (Total = 79)
clear (58), **succinct** (5), easy (4), straightforward (4), concise (2), [and 1 other with 1 occurrence] (Total = 74)	3 COMPREHENSIBILITY (15%) 38: 62	confusing (32), clearer (21), unclear (21), awkward (11), abstract (7), distracting (5), disjointed (4), difficult (2), disconnected (2), [and 15 others with 1 occurrence] (Total = 120)
true (21), **accurate** (13), careful (12), consistent (9), solid (6), valid (4), cogent (2), plausible (2), principled (2), systematic (2), [and 3 others with 1 occurrence] (Total = 76)	4 ACCURACY (12%) 49: 51	wrong (10), anecdotal (8), inconsistent (6), categorical (4), contentious (4), contradictory (4), inaccurate (3), unwarranted (3), bumpy (2), disingenuous (2), extreme (2), imprecise (2), loaded (2), relaxed (2), speculative (2), unfocused (2), [and 22 others with 1 occurrence] (Total = 80)
useful (72), valuable (15), important (14), helpful (10), substantial (10), practical (5), worthwhile (4), salient (3), crucial (2), meaningful (2), prominent (2), [and 1 other with 1 occurrence] (Total – 140)	5 IMPORTANCE (12%) 96: 4	meaningless (2), unhelpful (2), [and 2 others with 1 occurrence] (Total = 6)
sufficient (13), thorough (7), comprehensive (5), [and 1 other with 1 occurrence] (Total = 26)	6 SUFFICIENCY (10%) 20: 80	small (22), redundant (8), thin (8), briefer (5), fuller (5), narrow (5), underdeveloped (5), excessive (4), simplistic (4), brief (3), large (3), oversimplified (3), repetitive (3), lengthy (2), minimal (2), narrower (2), occasional (2), repetitious (2), superficial (2), unexplained (2), unsupported (2), [and 12 others with 1 occurrence] (Total = 106)
impressive (3), admirable (2), laudable (2), [and 2 others with 1 occurrence] (Total = 9)	7 PRAISEWORTHINESS (6%) 13: 87	disappointed (33), puzzled (7), disappointing (5), uncertain (3), frustrating (2), unsure (2), [and 11 others with 1 occurrence] (Total = 63)
sophisticated (6), insightful (4), intelligent (4), sensible (4), informed (3), **perceptive** (2), [and 3 others with 1 occurrence] (Total = 26)	8 PERCEPTIVENESS (3%) 81: 19	unaware (4), [and 2 others with 1 occurrence] (Total = 6)

Table 4: Classification of evaluative adjectives

(25) p. 5 'Knowledge built via the…' – *odd* sentence. (0032)

or that tense, capitalisation, expression, lexical choices, a collocation, or punctuation were odd in a section referred to, as in:

(26) I feel there's something *odd* about the tense used in 'the date in which the meeting will take place'. (0831)

COMPREHENSIBILITY adjectives (e.g. *clear*, *easy*, *lucid*, *straightforward*, and conversely *awkward*, *clearer*, *confusing*, *contorted*, *dense*, *difficult*, *obscure*, *incomprehensible*, *unclear*) are not only to do with comprehensibility, but also organisation and cohesion. The fact that these together constitute 15 per cent of the total suggests that the clarity of expression in the paper was valued highly by the reviewers. Entities most frequently judged to be (un)clear included: *account*, *analysis*, *argument*, *description*, *English*, *organisation*, *presentation*, and *style*. Other adjectives in this group (e.g. *concise*, *succinct*, and conversely *rambling*, *verbose*) suggest that economy of expression was also valued.

ACCURACY adjectives (12 per cent of the total) are to do with the truth or accuracy of entities (e.g. *accurate*, *true*, *valid*, and conversely *false*, *flawed*, *inaccurate*, *inexact*, *invalid*, *specious*, *spurious*, *questionable*, *untenable*, *untrue*, *wrong*), and whether they are logical and consistent (e.g. *consistent*, *logical*, *principled*, *systematic*, and conversely *inconsistent*). In addition, they seem to show a concern with whether entities (particularly statements and claims) are persuasive, based on valid evidence, and will stand up to scrutiny (e.g. *cogent*, *robust*, and conversely *anecdotal*, *contentious*, *extreme*, *overblown*, *speculative*, *unjustified*, *unwarranted*). Other adjectives in this group suggest that the reviewers valued care taken by authors (e.g. *careful*, and conversely *casual*, *cavalier*, *glib*, *sloppy*), particularly with respect to presentation and argument.

The journal *English for Specific Purposes* is expressly concerned with reporting research with outcomes for the practice of ESP, and papers are expected to make recommendations on implications for pedagogy. Many uses of IMPORTANCE adjectives (12 per cent of the total) evaluated the contribution that an entity is judged to make to readers or the discipline as a whole in this respect. For example:

(27) This paper reports an *important* and *valuable* study of written advice on the rhetorical organisation of theses, and of actual practice, which will be of interest to EAP practitioners. (0472)

In other cases, however, the importance of entities as elements of the reported research was evaluated:

(28) The specific examples of students work are *useful*… (0042)

SUFFICIENCY adjectives (10 per cent of the total) are mainly concerned with whether an entity is judged to be sufficient for a particular purpose (e.g. *comprehensive*, *exhaustive*, *sufficient*, *thorough*) or not. There may be too much of an entity (e.g. *briefer*, *excessive*, *lengthy*, *repetitive*) or too little (e.g. *brief*, *fuller*, *narrow*, *small*, *superficial*, *thin*). An entity may be unnecessary (e.g. *redundant*), lacking (e.g. *unasked*, *unexplained*, *unsupported*), or incomplete (e.g. *underdeveloped*, *unfinished*).

PRAISEWORTHINESS adjectives (6 per cent of the total) are to do with the effect that the quality of an entity is said to have on the reviewer or is projected will have on the wider readership. The reviewers indicated that they found entities *admirable*, *impressive*, *laudable*, *lovely*, or conversely *amusing*, *arrogant*, *disappointing*, *frustrating* or *worrying*. On other occasions they indicated that they were *impressed*, or conversely *disappointed*, *irritated*, *puzzled*, *uncertain* or *unsure* about or by an entity.

PERCEPTIVENESS adjectives (3 per cent of the total) are mainly to do with qualities the author is said, explicitly or implicitly, to have or to lack. Through what they write, authors show themselves to be, according to the reviewers, *insightful*, *intelligent*, *sophisticated*, etc., or conversely, *unaware*, *uninformed* or *vacuous*. The most common adjective in this category, *sophisticated* (and also *perceptive*) is used in the guidelines for authors.

6 Discussion and conclusions

The main findings from the two analyses can be summarised as follows. Four of the nine entity classes – PAPER, EXPRESSION, ANALYSIS and CLAIM – together comprised 80 per cent of all the entities evaluated (excluding those which had fewer than three occurrences in total), and each of these had around 20 per cent of the total. This suggests, then, that the main preoccupation of the reviewers was with entities within these classes. In general, that is: an overall judgement of the quality of a paper and its contribution to the discipline; the way in which the research is expressed and presented; the quality of the analysis and findings; and the validity of the author's interpretation of findings, including the claims made on the basis of them. In three of these entity classes there is a heavy preponderance of either positive or negative evaluation: PAPER entities are predominantly positive (75 per cent), while EXPRESSION and CLAIM entities are predominantly negative (70 per cent and 64 per cent respectively). In the ANALYSIS class the distinction is less clear (45 per cent positive).

Six of the eight groups of evaluative adjectives – INTEREST, SUITABILITY, COMPREHENSIBILITY, ACCURACY, SUFFICIENCY and IMPORTANCE – together comprised 90 per cent of the total number of evaluative adjectives used. From this perspective, then, the main preoccupations of the reviewers appears to be with

whether entities are appropriate for their purpose (for example, that the paper will interest the readership, that findings are important for the discipline, or that appropriate references are provided), whether entities are accurate and used in sufficient numbers, and that entities can be understood. The relative proportions of positive and negative adjectives in each of these dimensions suggest that the reviewers highlighted the *presence* of INTEREST, SUITABILITY and IMPORTANCE, but the *lack* of COMPREHENSIBILITY and SUFFICIENCY. No clear distinction is apparent within the ACCURACY group.

This paper has focused on a quantitative analysis of evaluated entities and evaluative criteria as expressed in adjectives attached to entities. It cannot be said with certainty that a high frequency of occurrence of a particular entity or a particular adjective represents a high level of importance attached by the reviewers to that entity or quality. However, it does suggest a high degree of attention paid to that entity or quality and this in turn suggests that the reviewers attached particular significance to it. With this caution in mind, this discussion will focus on two main questions: What do the findings suggest about (i) the entities that the reviewers valued highly (and, conversely, did not value highly)? and (ii) the relationship between the reviewer(s) and author(s)?

There was a clear concern among the reviewers that papers should make a contribution to the discipline, and that they should be interesting and be saying something that is original and important. This is suggested by the relatively high frequency of the number of evaluated entities in the PAPER class and the number of adjectival evaluative acts in the INTEREST dimension. It should be noted, however (as mentioned above), that many of the positive evaluations of the *paper* and its *contribution* are part of a pattern in which the reviewers said something positive at the beginning of the review before going on, in most cases, to make primarily critical comments in the remainder of the text or even to reject the paper. It may be that the high frequency of the adjectives *interesting* and *original* is the result of the reviewers genuinely finding papers interesting and original. However, another possibility is that it is the necessary consequence of the frequent use of this rhetorical pattern – saying something positive in a general way, before going on to offer more detailed criticism – which may well be part of an attempt to mitigate the discouraging effects of criticism in a document that will be read by the reviewer's peers.

The reviewers expressed their preference for papers to be written accurately and in an understandable way. This is suggested by the relatively high frequency of the number of evaluated entities in the EXPRESSION class and the number of evaluation acts in the COMPREHENSIBILITY dimension. Most reviewers took it upon themselves to point out language errors and infelicities, and often suggested corrections, even though they are not explicitly asked to do so in the guidelines. This might simply be in response to identifying places where they

had some difficulty understanding sections of papers. In addition, however, it might be seen to be a consequence of their seeking to maintain standards of expression or (to put it in perhaps a more constructive way) helping authors work towards a standard. It is worth remembering that the academic area of study of the reviewers investigated in this research is the English language, and this may result in their paying more attention to these matters than would, say, reviewers for journals in the sciences. In Gosden's (2001; 2003) research on a science journal, for example, most comments on English language were of a general type while those noted here tended to be quite detailed and specific.

The reviewers considered that claims made by authors should be justified and valid. This is suggested by the relatively high frequency of the number of evaluated entities in the CLAIM class and the number of evaluation acts in the ACCURACY dimension, and accords with Kourilová's (1998) finding that reviewers' criticisms of unjustified conclusions were most severe and Gosden's (2001; 2003) finding that 46.6 per cent of reviewers' comments were on discussion and claims. This was particularly clear in the relatively high frequency of negative evaluations of the entity *claim* (including *generalisation*, *statement*, and *assertion*). The following is typical:

(29) It is also unclear as to whether the claims made about academic writing are fully *appropriate*, given that… (0821)

Certain entities within entity classes also merit attention. There is either exclusively or predominantly negative evaluation of the entities *term* (all 18 instances negative) and *definition* (all 7) within the ANALYSIS class and *sample* (all 10), *evidence* (7 out of 10) and *corpus* (6 out of 7) within the EVIDENCE class. This suggests either that particular attention was paid by the reviewers to the correct use of terms, the forms of definitions, the size of samples, etc., or that these were particularly problematic areas for authors.

The categorisation of both the evaluated entities and the evaluative adjectives suggests that the reviewers take on multiple roles in the review process, and in different roles they are representative of different academic communities. As experienced researchers and representatives of the research community in general, they offer judgement on those entities which may be assessed objectively, such as the accuracy of findings, definitions, and the use of terminology, or the quantity of data necessary to meet the stated aims of the paper. As published authors themselves and representatives of the community of academic writers, judgement is made about the quality of language and clarity of expression. As experts in the particular field and representatives of the discipline of applied linguistics (or ESP in particular), they evaluate entities that allow only subjective judgement, such as the validity of claims or the originality of the material. As representatives of the readership of the journal

English for Specific Purposes, they make judgements on how interesting and comprehensible entities are likely to be. In this complexity of roles, they act as gatekeepers, only allowing access to publication when they believe that authors have satisfied the requirements (as the reviewers perceive them) of the different parts of the academic community.

On one hand, then, the reviewers represent the established 'core', preventing access to publication to those who do not attain the required standards of research quality or presentation. At the same time, however, the reviewers offered encouragement to authors that their paper includes something of merit, so that they will continue to revise their work. This may be manifested in the evidence presented here in the high proportion of positive evaluations of PAPER entities. Similarly, the large number of evaluations of EXPRESSION entities may be seen as a sign of the reviewers guiding authors towards improved presentation of their research. While most evaluations are of entities that are part of the research process or the written text, some relate to the authors themselves. In the KNOWLEDGE entity class, the authors' *awareness*, *understanding*, etc. are evaluated, and in the PERCEPTIVENESS group of adjectives authors are directly or indirectly said to be *astute*, *informed*, etc. Reviews are written for editors and are not addressed directly to authors, but reviewers know that reviews will be read usually in an unchanged form by authors, and it seems likely that such evaluations are addressed more to authors to offer praise (most of these evaluative acts are positive) than to inform editors about the qualities of authors. Such evaluation, then, would seem to be a reflection of the reviewer's role in offering encouragement, as, too, may be the mitigation of criticism noted above in the pattern of 'saying something positive before saying something negative'. Further evidence of the support offered by the reviewers may be found from the results presented in Table 2. The fact that reviews recommending that papers should be revised and resubmitted were, on average, over twice the length of those recommending publication and nearly twice the length of those rejecting papers suggests that the reviewers offered guidance in how papers should be revised, either directly through suggestions or indirectly through identifying weaknesses. Further investigation would be needed, of course, to substantiate this possibility.

As well as taking on multiple roles, it should be remembered that reviewers write for multiple audiences. In addition to writing for the journal editor and authors, reviewers for *English for Specific Purposes* are (in most cases) aware that their reviews are shared with other reviewers of the same paper, albeit anonymously. Any impact of this on the form of evaluation would be hard to specify precisely, but it could be, for example, that more extreme evaluations are avoided so as to prevent substantial disagreement with the opinions of another reviewer. Offering severe criticism of a paper or some aspect of it while another

reviewer has provided fulsome praise may be embarrassing even if only the editor knows the identities of the two reviewers.

This work has contributed a perspective on the language of peer reviews from the social sciences to add to those reported for science journals. However, it is offered very much as a preliminary study, among the purposes of which is to identify productive avenues of further research. The paper will close by identifying some of these. First, the study has focused only on the most transparent form of evaluation in the reviews: through the use of evaluative adjectives. Other linguistic means are used to express evaluation (evaluative nouns, modal verbs, etc.), and these need to be investigated in order to provide a more complete picture. Second, the focus of attention has been the *product* of review rather than the process of reviewing itself. To better understand evaluation in peer reviews, information needs to be gathered from reviewers themselves on their motivations for selecting entities for evaluation and the criteria by which they judge these entities, and also on the influences on the form of reviews. The reviewers in this study have all read reviews written by others – reviews assessing their own papers submitted for publication, and also reviews shared by the editors of *English for Specific Purposes* (see above). They are likely, then, to have some sense of the organisation and content of the 'peer review genre'. Their perceptions of this and how these influence their own review writing would seem to be fruitful areas of further exploration. Similarly, it is only through discussions with reviewers that the impact of *English for Specific Purposes*' guidelines for reviewers might become apparent. Third, no attempt has been made in this study to differentiate between the papers in the corpus according to the type of research reported (whether, for example, it is quantitative or qualitative), or between the authors or the reviewers (whether, for example, they are native or non-native speakers of English). Evaluation might vary depending on characteristics such as these. Finally, it has been argued that as the reviewers are representatives of the wider academic community, so the values reflected in their evaluations are also those of that community. However, it remains to be investigated how wide this community is: are the values displayed in this study also found in the peer reviews of other journals in the field of applied linguistics, and in those for journals in other disciplines?

References

Austin, J. L. (1962) *How to do Things with Words*. Oxford: Clarendon Press.

Bazerman, C. (1988) *Shaping Written Knowledge*. Madison: University of Wisconsin Press.

Benfield, J. R. and Howard, K. M. (2000) The language of science. *European Journal of Cardiothoracic Surgery* 18: 642–8.

Berkenkotter, C. and Huckin, T. (1995) *Genre Knowledge in Disciplinary Communication*. Hillsdale, NJ: Lawrence Erlbaum.

Bordage, G. and Caelleigh, A. S. (2001) A tool for reviewers: review criteria for research manuscripts. *Academic Medicine* 76: 904–8.

Burrough-Boenisch, J. (2003) Shapers of published NNS research articles. *Journal of Second Language Writing* 12: 223–43.

Canagarajah, A. S. (1996) 'Non-discursive' requirements in academic publishing, material resources of periphery scholars, and the politics of knowledge production. *Written Communication* 13(4): 435–72.

Canagarajah, A. S. (2002) *A Geopolitics of Academic Writing*. Pittsburgh: University of Pittsburgh Press.

Canagarajah, A. S. (2003) A somewhat legitimate and very peripheral participation. In C. P. Casanave and S. Vandrick (eds) *Writing for Scholarly Publication*: *behind the scenes in language education* 197–210. Mahwah, NJ: Lawrence Erlbaum Associates.

Casanave, C. P. and Vandrick, S. (2003a) Introduction: Issues in writing for publication. In C. P. Casanave and S. Vandrick (eds) *Writing for Scholarly Publication*: *behind the scenes in language education* 1–13. Mahwah, NJ: Lawrence Erlbaum Associates.

Casanave, C. P. and Vandrick, S. (2003b) *Writing for Scholarly Publication*: *behind the scenes in language education*. Mahwah, NJ: Lawrence Erlbaum Associates.

Chilton, S. (1999) The good reviewer. *Academe* 85 (6): 54–5. Accessed at http://www.d.umn.edu/~schilton/Articles/Reviewer.html on 21st December 2004.

Cronin, B. (1984) *The Citation Process*. London: Taylor Graham.

Curry, M. J. and Lillis, T. (2005) Multilingual scholars and the imperative to publish in English: negotiating interests, demands and rewards. *TESOL Quarterly* 38: 663-88.

Ernst, E. and Resch, K. L. (1994) Reviewer bias: a blinded experimental study. *Journal of Laboratory and Clinical Medicine* 124: 178–82.

Flowerdew, J. (1999a) Problems in writing for scholarly publication in English: the case of Hong Kong. *Journal of Second Language Writing* 8(3): 243–64.

Flowerdew, J. (1999b) Writing for scholarly publication in English: the case of Hong Kong. *Journal of Second Language Writing*, 8(2): 123–145.

Flowerdew, J. (2000) Discourse community, legitimate peripheral participation, and the non-native English-speaking scholar. *TESOL Quarterly* 34(1): 127–50.

Flowerdew, J. (2001) Attitudes of journal editors to non-native-speaker contributions: an interview study. *TESOL Quarterly* 35(1): 121–50.

Fox, R. (2003) Centre and periphery revisited. The structures of European science, 1750–1914. Introduction. *Revue de la Maison Française d'Oxford* 1, 2. Accessed at http://www.mfo.ac.uk/Publications/Revue%20Fox/introduction.htm on 18th March 2005.

Gilbert, G. N. (1976) The transformation of research findings into scientific knowledge. *Social Studies of Science* 6: 281–306.

Gilbert, G. N. and Mulkay, M. (1982) Warranting scientific belief. *Social Studies of Science* 12: 383–408.

Godlee, F., Gale, C. and Martyn, C. (1998) Effect on the quality of peer review of blinding reviewers and asking them to sign their reports. *The Journal of the American Medical Association* 280(3): 237–40.

Goldbeck-Wood, S. (1998) What makes a good reviewer of manuscripts? *British Medical Journal* 316: 86.

Gosden, H. (2001) 'Thank you for your critical comments and helpful suggestions': compliance and conflict in authors' replies to referees' comments in peer reviews of scientific research papers. *IBÉRICA* 3: 3–17.

Gosden, H. (2002) Thematic content in peer reviews of scientific papers. *Seminários de Linguística* 5.

Gosden, H. (2003) 'Why not give us the full story?': functions of referees' comments in peer reviews of scientific research papers. *Journal of English for Academic Purposes* 2(2): 87–102.

Graddol, D. (1997) *The Future of English*? London: British Council.

Hyland, K. (2000) *Disciplinary Discourses*: *social interactions in academic writing*. London: Longman.

Hunston, S. (1993) Evaluation and ideology in scientific writing. In M. Ghadessy (ed.) *Register Analysis*: *theory and practice* 57–73. London: Pinter.

Jefferson, T., Wager, E. and Davidoff, F. (2002) Measuring the quality of editorial peer review. *Journal of the American Medical Association* 287(21): 2786–90.

Kourilová, M. (1998) Communicative characteristics of reviews of scientific paper written by non-native users of English. *Endocrine Regulations* 32: 107–14.

Kuhn, T. (1970) *The Structure of Scientific Revolutions.* (Second edition) Chicago: University of Chicago Press.

Latour, B. and Woolgar, S. (1979) *Laboratory Life*: *the social construction of scientific facts*. Beverly Hills: Sage.

Lave, J. and Wenger, E. (1991) *Situated Learning*: *legitimate peripheral participation*. Cambridge: Cambridge University Press.

McNutt, R. A., Evans, A. T., Fletcher, R. H., et al. (1990) The effects of blinding on the quality of peer review: a randomised trial. *Journal of the American Medical Association* 263: 1371–6.

Myers, G. (1990) *Writing Biology*: *texts in the construction of scientific knowledge*. Madison: University of Wisconsin Press.

Nylenna, M., Riis, P. and Karlsson, Y. (1994) Multiple blinded reviews of the same two manuscripts: effect of referee characteristics and publication language. *Journal of the American Medical Association* 272(2): 149–51.

Pennycook, A. (1994) *The Cultural Politics of English as an International Language*. London: Longman.

Phillipson, R. (1992) *Linguistic Imperialism*. Oxford: Oxford University Press.

Rowland, F. (2002) *The Peer Review Process*. A report to the JISC Scholarly Communication Group. Accessed at http://www.jisc.ac.uk/uploaded_documents/rowland.pdf on 14th December 2004.

Scott, M. (1996) *Wordsmith Tools*. Oxford: Oxford University Press.

Searle, J. R. (1969) *Speech Acts*. London: Cambridge University Press.

Swales, J. (1981) *Aspects of Article Introductions*. (Aston ESP Research Reports 1.) Birmingham: Language Studies Unit, University of Aston.

Swales, J. (1996) Occluded genres in the academy: the case of the submission letter. In E. Ventola and A. Mauranen (eds) *Academic Writing*: *intercultural and textual issues* 45–58. Amsterdam: John Benjamins.

Thetela, P. (1997) Evaluated entities and parameters of value in academic research articles. *English for Specific Purposes* 16: 101–18.

Thompson, G. and Hunston, S. (2000) Evaluation: an introduction. In S. Hunston and G. Thompson (eds) *Evaluation in Text*: *authorial stance and the construction of discourse* 1–27. Oxford: Oxford University Press.

van Rooyen, S., Godlee, F., Evans, S., Black, N. and Smith, R. (1999) Effect of open peer review on quality of reviews and on reviewers' recommendations: a randomised trial. *British Medical Journal* 318: 23–7.

Walsh, E., Rooney, M., Appleby, L. and Wilkinson, G. (2000) Open peer review: a randomised controlled trial. *The British Journal of Psychiatry* 176: 47–51.

Wood, D. and Hurst, P. (2000) Online peer review: perceptions in the biological sciences. *Learned Publishing* 13(2): 95–100.