

## LOOKING AT CITATIONS: USING CORPORA IN ENGLISH FOR ACADEMIC PURPOSES

**Paul Thompson**

Reading University

**Chris Tribble**

King's College London University & Reading University

### ABSTRACT

Appropriate reference to other texts is an essential feature of most academic writing, and we should expect courses in academic writing to sensitize students to the choices that are available to them when they decide to refer to other texts. A brief review of popular EAP writing textbooks finds, however, that attention is given mainly to surface features of citation, focusing on quotation, summary, and paraphrase.

Analysis of a purpose-built corpus of academic text can reveal much about what writers actually do, and can also generate rich speculation on why writers do what they do. Extending Swales' (1990) division of citation forms into integral or non-integral, we present a classification scheme and the results of applying this scheme to the coding of academic texts in a corpus. The texts are doctoral theses, written in two departments: Agricultural Botany and Agricultural Economics. The results lead into a comparison of the citation practices of writers in different disciplines and the different rhetorical practices of these disciplines. Comparison with Hyland (1999), which looks at citation types in research articles, also indicates differences between genres.

We then look at examples of EAP student writing and apply the same analysis to these texts. The results show that the novice writers use a limited range of citation types, and we suggest that teaching should focus on extending the range of choices available to students. Lastly, we introduce a number of class activities in which students conduct their own analyses of citation practices in small corpora, to develop genre awareness, and we evaluate these activities.

---

### INTRODUCTION

The growing interest in the application of corpus tools in language education, and the spread of "data-driven learning" (Tim John's coinage) is evidenced by the papers in this edition of *Language Learning and Technology*, and in recent publications (e.g., Burnard & McEnery, 2000). We will not, therefore, review the development of classroom concordancing in this short article or argue for its relevance to English language teaching, but will look immediately at an area of possible application for specific corpora.<sup>1</sup>

In this article, we report work that has used corpora to research a particular aspect of academic writing (citation practices, across the disciplines), how current ELT materials address the language features that were the focus of this research, and how corpus tools can be used to supplement published materials to give learners in EAP writing classes opportunities to extend their understanding of this central aspect of academic discourse.

## CORPUS-BASED RESEARCH INTO CITATION PRACTICES

Making references to the literature is an essential part of most academic writing, and it is also a source of considerable difficulty for most novice writers (Borg, 2000; Campbell, 1990). Some of the reasons that academic writers are expected to make references are to integrate the ideas of others into their arguments, to indicate what is known about the subject of study already, or to point out the weaknesses in others' arguments, aligning themselves with a particular camp/school/grouping. Novice writers may face problems because they are not at the appropriate stage of cognitive or intellectual development (Britton, Burgess, Martin, McLeod, & Rosen, 1975; Pennycook, 1996), or because of cultural factors (Connor, 1996; Fox, 1994). Failure to acknowledge the source of ideas can lead to charges of plagiarism, whereas inexpert phrasing of reporting statements can lead to confused or misleading indication of both the writer's, and the cited author's, stance (Groom, 2000).

Swales (1981, 1986, 1990) has pioneered the study of citation analysis from an applied linguistic perspective. He created clear formal distinctions between non-integral and integral citation forms: The former are citations that are *outside* the sentence, usually placed within brackets, and which play no explicit grammatical role in the sentence, while the latter are those that play an explicit grammatical role within a sentence. The citation at the beginning of this paragraph is an integral citation. He also used the terms "short" and "extensive," to describe citations that are at a single sentence level and those that encompass more than one sentence. These distinctions provide useful starting points but they do not provide insights that will help student writers understand which citation type to use in which context. Alongside Swales' work, there has been substantial research into the correlation of verb tense and voice in reporting verbs with function (most notably Shaw, 1992, but also worthy of mention are Hanania & Akhtar, 1985, and Malcolm, 1987).

Analysis of academic text corpora has the potential to inform our knowledge about the different forms and functions of citations in academic writing. Pickard (1995) used a small corpus of applied linguistics articles to investigate the citation practices of "expert" writers. On the premise that novice writers tend to overuse particular items in their references, such as "say," she investigated citation practices in the corpus to find out what expert writers do. Using concordancing software, she was able to produce statistical information to identify preferences among her writers for *integral* or for *non-integral* citation forms, and to identify the different grammatical forms of integral citations (subject, agent, genitive noun phrase, etc.). This was a useful preliminary study. The limitations were that the corpus was small, and there was little discussion of the reasons why writers choose one form rather than any other; the categories are based on syntactic distinctions rather than functional. More importantly, however, it is not clear whether her discoveries about the practices of a small number of applied linguistics writers can be generalized to "expert" writers across all the disciplines. It seems likely that writers in different disciplines follow different rhetorical conventions and have different preferences.

Two recent studies of citation practices in academic texts that test this assumption are Hyland (1999) and Thompson (2000). These two studies were based on the analysis of more substantially sized corpora, each investigating a different genre of academic writing. Hyland looked at citations in a corpus of 80 research articles, composed of 10 journal articles from different disciplines (see [Table 1](#) below for details), while Thompson (2000) examined differences in citation practices in a corpus of doctoral theses. The latter corpus contains 16 theses written in two departments at the University of Reading, 8 theses from the Department of Agricultural Botany, and 8 from the Department of Agricultural and Food Economics.

Table 1. Number of Citations in Hyland (1999) and Thompson (2000) Corpora

<b>Discipline</b>	<b>Av. per paper</b>	<b>per 1,000 words</b>
Mechanical Engineering	27.5	7.3
Physics	24.8	7.4
Electronic Engineering	42.8	8.4
Marketing	94.9	10.1
Philosophy	85.2	10.8
Applied Linguistics	75.3	10.8
Sociology	104.0	12.5
Biology	82.7	15.5
	<b>Av. per thesis</b>	
Agricultural Botany	248.8	9.04
Agricultural Economics	333.5	5.25

Both studies investigated variation in practice in disciplinary discourses and made use of frequency and concordance data to investigate dispersion, frequency, and patterning across large quantities of text. Table 1 shows the figures for instances of citation in the two corpora, with the middle column showing the average number of citations per text, and the right column showing the average number of citations per 1,000 words of running text. The lower density of citations amongst the science and technology articles (7.3-8.4) contrasted with higher incidence among the social science articles (10.1-12.5) while Biology stood out as exceptional with 15.5. Hyland postulated a difference in practice here between "hard" and "soft" disciplines, using terminology drawn from Becher (1989), and speculated that Biology stood out from the other sciences because it is a relatively new discipline. The distinction between "hard" and "soft" disciplines may, however, prove to be reductive; Becher himself prefers a multi-dimensional model with added axes of "applied" and "pure," "rural" and "urban." The fact that the Biology texts are so markedly different from the Physics and Engineering texts is evidence that the simple distinction between "hard" and "soft" is inadequate.

As can be seen in Table 1, the density of citations in the doctoral theses is much lower. If we presume that the Agricultural Botany theses should be roughly comparable to the Biology articles, the density is approximately three fifths lower. And while there is no easy comparison between the Agricultural Economics and any of the disciplines in Hyland's study, the figure of 5.25 is substantially lower than any of the figures for the research articles. It can be seen, therefore, that the two genres are marked by different degrees of use of citations. One explanation for this is that the types of texts produced in these two genres are of different lengths: Articles usually average between 2,000 and 5,000 words, while in Thompson's study, the average length of an Agricultural Botany thesis was 31,000, and the average length of an Agricultural Economics thesis was 63,000. As articles are shorter texts, there is presumably a need for a more condensed style of writing.

Table 2 shows the relative percentages of the two types of citation, integral and non-integral, in Hyland (1999) and Thompson (2000). These figures show firstly that there is considerable variation in citation practice between the different disciplines, with Philosophy being the only discipline that prefers the integral form over the non-integral, greater emphasis being placed on the arguments of different individuals. Secondly, it is interesting to note that in the case of the Agricultural Economics theses writers, the integral type was also preferred. Although no direct comparison can be made between Agricultural Economics and the disciplines in Hyland's study, one would not expect Agricultural Economics to be closest to Philosophy. A more plausible explanation is that citation practices in the two genres are different: Thesis writers in Agricultural Economics make greater use of integral citations for

reasons that become clear from closer reading of the texts. One obvious point is that the length of texts in the two genres is markedly different: the articles in Hyland's corpus range from 3 to 31 pages in length, whereas the Agricultural Economics theses in Thompson's corpus are around 200 pages long. In long texts, such as the Agricultural Economics theses, or in book length treatments of research, there is a higher likelihood that references to leading researchers in the field will be elaborated and give greater prominence to the author(s).<sup>2</sup>

Table 2. Ratios of Non-Integral to Integral Citations by Discipline in Hyland (1999) and Thompson (2000)

<b>Discipline</b>	<b>Non-integral</b>	<b>Integral</b>
Biology	90.2	9.8
Electronic Engineering	84.3	15.7
Physics	83.1	16.9
Mechanical Engineering	71.3	28.7
Marketing	70.3	29.7
Applied Linguistics	65.6	34.4
Sociology	64.6	35.4
Philosophy	35.4	64.6
<b>Doctoral theses</b>		
Agricultural Botany	66.5	33.5
Agricultural Economics	38.1	61.9

Table 3 below shows the percentage of citations in the two corpora that incorporate direct quotation from the source text. It is clear from these figures that quotation is a relatively common feature in the social science and humanities texts but that it is scarcely used in the science texts. Where quotation is used in the science texts, (viz. the 0.8% figure in the Agricultural Botany column), the citation is a definition, while many of the Agricultural Economics quotations are evaluative comments.

Table 3. Sample Percentages of Citations in Two Corpora That Include Direct Quotation

Articles (Hyland, 1999)				Doctoral theses (Thompson, 2000)	
Biology	Electronic Engineering	Sociology	Applied Linguistics	Agricultural Botany	Agricultural Economics
0	0	13	10	0.8	8

The statistics reported here suggest that there are clear divergences in the citation practices of writers in different disciplines, and also between genres of academic writing.

The level of analysis at this stage, however, restricts the kinds of questions that one can ask, and it is necessary to develop a more sensitive set of categorisations. In the next section, we describe a set of categories drawn from Thompson (2000), and we then proceed to pose further questions.

## NON-INTEGRAL CITATION

### Source

Non-integral citations perform a range of functions. The first function is to attribute a proposition to another author. The proposition might be a statement of what is known to be true, such as in the factive report of findings in other research, or the attribution of an idea to another, as in this example:

Citation is central ... because it can provide justification for arguments (Gilbert, 1976)

The citation provides evidence for a proposition which can remain unchallenged if the writer is in agreement with it, or can be countered by the ensuing argument. Let us call this type of citation *source* because it indicates where the idea comes from.

### Identification

The second type of non-integral citation identifies an agent within the sentence it refers to. An example of this is

A simulation model has therefore been developed to incorporate all the important features in the population dynamics (Potts, 1980)<sup>3</sup>

where the information within the parentheses identifies the author of the study referred to. Instead of including the name of the author within the sentence ("Potts [1980] has developed..." or "A simulation model has been developed by Potts [1980]..."), the writer has chosen to focus attention on the information (Weissberg & Buker, 1990, differentiate between author- and information-prominent citations).

### Reference

This type of citation is usually signalled by the inclusion of the directive "see" as in

DFID has changed its policy recently with regard to ELT (see DFID, 1998).

This type of citation is often similar to a source citation in that it can provide support for the proposition made, but it also functions as a shorthand device: Rather than provide the information in the present text, the writer refers the reader to another text. This type is particularly common in reference to procedures or to detailed proofs of arguments which are considered too lengthy to be repeated.

### Origin

An example of this type is

The software package used was Wordsmith Tools (Scott, 1996).

Where Source citations attribute a proposition to a source, Origin citations indicate the originator of a concept or a product - in this case the creator of the Wordsmith Tools programme.<sup>4</sup>

## INTEGRAL CITATIONS

A clear distinction can be made between integral citations which control a lexical verb (*Verb controlling*) and those that do not (*Naming*). A third type is the reference to a person that is not a full citation -- this has been called a *Non-citation* form.

### Verb Controlling

The citation acts as the agent that controls a verb, in active or passive voice, as in

Davis and Olson (1985) define a management information system more precisely as...

## **Naming**

In Naming citations, the citation is a noun phrase or a part of a noun phrase. The distinction here is primarily grammatical but the form also implies a reification, such as when the noun phrase signifies a text, rather than a human agent:

Typical price elasticities of demand for poultry products in Canada, Germany and the UK are shown in Harling and Thompson (1983)

Another example of reification is when the naming citation identifies a particular equation, method, formulation or similar construct with individual researchers, as in

In this paper, the management information system (MIS) definition of Davis and Olson (1985) has been used.

An alternative type of naming citation is that which refers generally to the work or findings of particular researchers:

Work by Samuel and East (1990) demonstrated that variety and seed rate had considerable effects on yield and quality aspects

In this case, the naming citation is similar to a verb-controlling citation in that it reports work done by particular researchers.

## **Non-citation**

There is a reference to another writer but the name is given without a year reference. It is most commonly used when the reference has been supplied earlier in the text and the writer does not want to repeat it. For example

The "classical" form of the disease, described by Marek, causes significant mortality losses.

However, instances where a person was invoked through reference to the thinking associated with them in general, rather than with reference to a specific work or set of works (for example, "Marxist" or "Darwinian") are not included.

## **FURTHER EXPLORATION**

Employing these categories, it is possible to explore a number of questions about the theses examined in Thompson (2000):

**Q1. Are there differences in the types of non-integral, or integral, citations used by writers in different disciplines?**

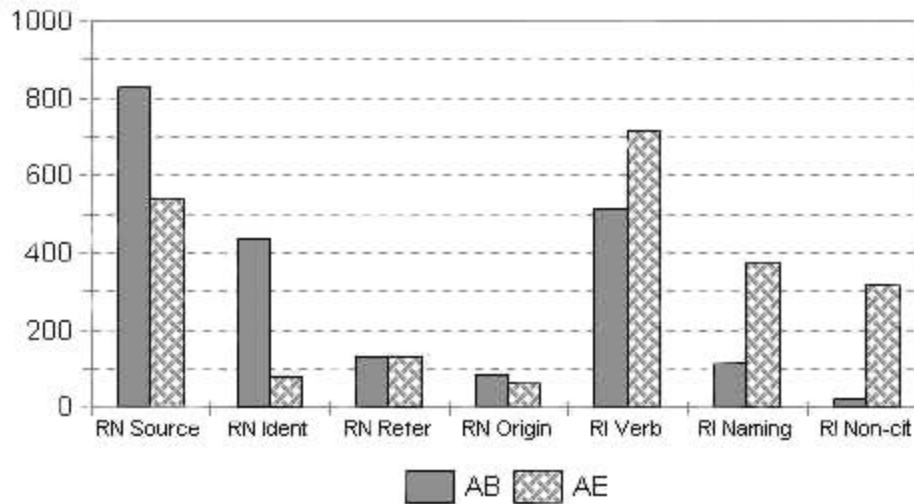


Figure 1. Proportion of citation types used in the two disciplines

As shown in [figure 1](#), writers in Agricultural Botany use the non-integral Source and Ident types much more frequently, while the Agricultural Economists make far greater use of integral Naming citations (reasons for which become apparent in [Q4](#) below) and also make more mentions of names without giving full citation information.

**Q2. Are there differences in the practices of writers within the same discipline?**

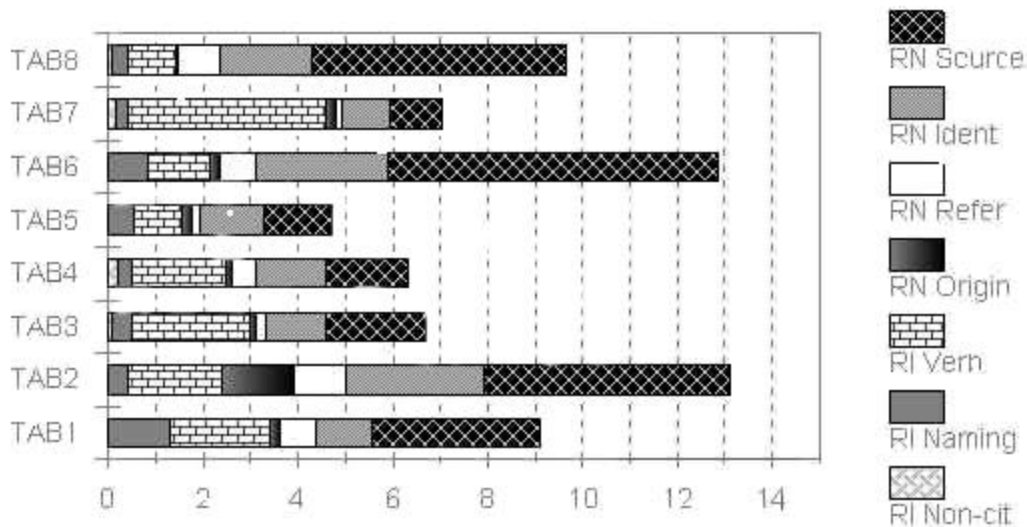


Figure 2. The average number of different citation types per 1,000 words of text found in the eight Agricultural Botany theses

As can be seen in [Figure 2](#), the density of citations in the individual Agricultural Botany theses varies from just under 5 per 1,000 words (TAB5) to around 13 (TAB2 and TAB6). TAB7 uses Verb-controlling citation types far more than any of the other writers, and far fewer non-integral citation types. Examination of this thesis reveals that the writer makes frequent reference to individual studies and compares their findings to his own experiments (X found this, and Y reported this. My findings were ...). TAB 6, by contrast, uses predominantly non-integral citation forms, and prefers to make information prominent through use of the Identification citation rather than the integral Verb controlling type. TAB6 is a report of a laboratory-based investigation of innovative techniques for isolation of vacuoles, and



therefore the emphasis is on the techniques, and the subject of study, that is, the vacuoles. Different writers within one discipline, then, take different approaches to research, and their rhetorical choices are, to a degree, determined by the nature of the research that they conduct.

### Q3. Are different types of citation used in different rhetorical sections?

In the Agricultural Botany theses, it was possible to divide the texts into four types of rhetorical section, following the conventions that are common in most scientific reports: Introduction, Methods, Results, Discussion. As can be seen in Table 4, there is considerable variation in the different sections of the theses, with relatively low use of citations in the Methods and Results sections of the thesis, and a markedly different set of citation types in the case of the Methods sections. To understand these variations, it is helpful to think of the hourglass model proposed by Hill, Soppelsa, and West (1982): the Introduction and Discussion sections of an article take a broad view, relating what is known in the field at large, while the Methods and Results sections are narrow, focussing on the research itself. While the Introduction and Discussion sections contain many references to other studies to establish the current state of knowledge and where the current report fits in, the Methods section contains mainly references to the methods and techniques of others.

Table 4. Citation Types in Different Rhetorical Sections of AB Theses

Section	Density (per 1,000 words)	Most common types of citation
Introduction	15.6	Source, Identification, Verb controlling
Methods	2.3	Refer, Origin, Naming
Results	2.4	Source (52%)
Discussion	10.1	Source, Identification, Verb controlling

This data shows that there is, then, variation in the density and type of citations used in different rhetorical sections of a thesis, and similar variation has been found across rhetorical sections in Physics, Chemistry and Biology masters' theses (Hanania & Akhtar, 1985).

Table 5. The Number of Occurrences of Naming Citations in the Two Disciplinary Groupings

RI Naming	Total occurrences
AB	116
AE	484

### Q4. Are there differences in patterns of language around particular citation types?

Close inspection of the different kinds of Naming citation in the theses revealed interesting differences in the discourses of the two disciplines. Firstly, in terms of simple frequency, it can be seen from Table 5 that this citation type is much more commonly used (by more than four times) in the Agricultural Economics texts. In order to find out why this might be the case, concordance lines of the Naming citation type were examined. It was observed that certain patterns were regularly used, such as the three shown in Table 6.

Table 6. The Number of Occurrences of a Pattern in the Thesis Corpus of Preposition + Naming Citation

Agricultural Botany		Agricultural Economics
3 (12)*	...in X (1991)	58
37 (154)	...of X (1991)	70
25 (104)	...by X (1991)	29

\* In the middle column, the figure in brackets shows an adjusted figure which would make the amount equivalent to the figure in the right column ( $n \times 484/116$ ).



The pattern "in X (1991)" is clearly much more commonly used in the Agricultural Economics theses. The use of the preposition "in" indicates that the citation is a reference to a book, and this is supported by the examples given in Table 7. In the Agricultural Botany theses on the other hand, "of" and "by" are more commonly used and these tend to refer to the research actions, findings, methods, and techniques of other researchers. Where Agricultural Economics thesis writers use "of," it is noticeable that this also includes discourse nouns, such as views and suggestions, which are not found in the Agricultural Botany texts. The Agricultural Economics writers, therefore, appear to be concerned with the texts and concepts of others, while the Agricultural Botany writers make reference to the research activities and techniques of other scientists.

Table 7. Frequent Patterns Involving in "in," "by," and "of" in the theses

<i>Agricultural Botany</i>		<i>Agricultural Economics</i>	
findings	of X	listed	in X
work		found	
method		extrapolated	
technique		quoted	
work	by X	views	of X
experiments		suggestions	
studies		model	
		function	
		exposition	

### REASONS FOR VARIATION

We have seen from the quantitative data that there are substantial differences in citation practices between disciplines and between genres. The types of research work undertaken, the epistemological bases upon which this research is founded, the conventions of the discipline, and the purposes for which texts are created all influence the forms of citation made. Looking at citation from a micro-perspective, however, one might naturally ask, "What is it that leads a writer to choose one citation form over another?" Why, for example, did we choose, earlier in this paper, to write "Two recent studies of citation practices in academic texts that test this assumption are Hyland (1999) and Thompson (2000)," rather than "Hyland (1999) and Thompson (2000) are two recent studies of citation practices in academic texts that test this assumption"?

Our reason in this case was that we wanted to place the noun phrase beginning "two recent studies" in theme position within the sentence. Shaw (1992) has observed that this is commonly the factor that determines voice (active/passive) in reporting verbs in sentence construction. The choice between using a non-integral identification type ("A simulation model has therefore been developed to incorporate all the important features in the population dynamics [Potts 1980]") instead of a Verb-controlling type ("A simulation model has been developed by Potts [1980] ...") is often governed by decisions as to how much prominence to give to the people involved (cf. Weissberg & Buker, 1990). To a certain extent, disciplinary convention plays a part here; it is conventional in scientific writing to de-emphasize the role of the researchers, particularly in controlled experiments, where the claim is that the human factor is not consequential (Dr. Philip John, School of Plant Sciences, University of Reading, personal communication).

## WHAT DO EAP TEXTBOOKS SAY ABOUT CITATIONS?

In the previous sections we have outlined a number of research findings regarding both the kinds of citations that are used in "expert performances" (Bazerman, 1994, p. 131) and the reasons for their use. Our next task is to review what kinds of advice or models are provided in published materials for EAP students, and to assess the extent to which these might need complementing. Three widely used EAP course books were selected for this purpose: Jordan (1992), Trzeciak and Mackay (1994), and Swales and Feak (1994). In summary, the course books provided surprisingly little advice or guidance to learners.

Jordan (1992) offers little explicit advice and depends mainly on quotations to provide models for learners to work from. However, as Jordan only exemplified three kinds of citation it cannot be considered a sufficient treatment of the subject:

1. *non-integral* "... (Seers, 1979, pp. 27-28) a further dimension is added - 'development now implies, *inter alia*, that...'"
2. *integral - naming* "... For Seers, 'Development is inevitably a normative concept' ... (Seers, 1972, p 22)"
3. *integral - verb controlling* "... Hicks and Streeten (1979, p 568) identify and review four different approaches..."

Similarly, Trzeciak and Mackay (1994) comment on only three kinds of citation:

1. *integral - verb controlling* ... Reporting using paraphrase
2. *non-integral - identification* ... Reference to source
3. *integral - other* ... Direct quotation...

But again, they offer little in the way of clear guidance to the apprentice writer and do not draw their attention to disciplinary differences.

Swales and Feak (1994) give a relatively fuller range of advice and examples, and discuss the contrast between non-integral/integral and footnote styles. However, they make no comment on the implications of using contrasting forms, and fall back on references to APA and MLA style guides. They do, however, usefully comment on the role of citation in abstracts.

In conclusion, it is possible to say that little explicit advice is given in major teaching materials on how to manage citations in specific disciplines. Instead, there is an emphasis on summary, paraphrase, and quotation, and on a small set of the mechanical features associated with citation. How then can students learn more about citation practices in their own subject area?

## USING MICRO-CORPORA TO COMPLEMENT EAP WRITING PROGRAMMES

Arguments have been made for the development of micro-corpora as resources for use in EAP programmes (Hyland, 2000; Tribble, 2001), and a corpus-informed approach appears to have much to recommend itself so long as relevant data are available. The need for such support is reinforced when student use of citations is investigated. In the preparation of this paper we reviewed a small collection of student assignments written at Reading university,<sup>5</sup> and identified the following problems:

- Lack of variety of citation types within single texts (e.g., the repeated use of "According to...")
- Lack of linguistic variety + inappropriate selection of verb (e.g., inappropriate use of "claims")
- Absence of certain categories (e.g., Non-integral reference)
- Over-use of non-citational references to authors / authorities

These findings (supported by extensive experience of teaching EAP students) indicate that two kinds of resource will be of benefit to learners: firstly, a collection of their own writing, or the writing of their peers -- a "learner corpus" (Granger, 1998), and, secondly, a collection of examples of writing from the target discourse community (e.g., research articles/dissertations, etc., from the students' own field of study), or texts as closely analogous to this kind of writing as possible (e.g., student examination scripts -- which are notoriously difficult to get hold of). While the collection of such data banks used to be difficult and time consuming, with the use of word-processors by students and the growing availability of electronic texts from the WWW or low cost scanners, and accurate optical character recognition (OCR) programs,<sup>6</sup> these restrictions no longer really apply.

With appropriate text resources to hand, it is relatively easy for teachers and students to begin a systematic investigation of citation practice in genres that are relevant to their own needs or interests. This need not require the use of a concordancing program; setting the search function in a word-processor such as Microsoft Word® to look for "(19" or "(20" with the "Find whole words only" un-checked will provide rapid access to the dated citations in a text, as will a search for a list of names based on the bibliography in an article. Obviously, more powerful searching and analysis of the results will be possible with a dedicated concordancing program.<sup>7</sup>

An appropriate procedure will be

- Stage 1 learners are introduced to a range of citation forms appropriate to their level of study
- Stage 2 learners investigate actual practice in relevant texts, reporting back on the form and purpose of citations they identify
- Stage 3 learners investigate the practices of their peers in writing assignments
- Stage 4 learners review their own writing and revise in the light of these investigations.

As an example of how such a procedure can be used in an EAP programme we have drawn on the British National Corpus<sup>8</sup> -- making use of Dave Lee's BNC Index (see [Lee article](#) in this issue) to make a micro-corpus of 22 extracts from one academic journal -- *Language and Literature*.<sup>9</sup> The assumption in this case has been that the texts will be of interest to post-graduate humanities EAP students who are (a) interested in extending the ways in which they word citations, and who (b) wish to ensure that they are writing in a way that is appropriate for their field of study. It is possible for a teacher to use the four stage procedure outlined above to develop a set of learning materials which will achieve this end.

In Stage 1, students will be familiarised with the citation categories we have discussed in this article. In Stage 2, they will work in different groups to complete a task such as the one given below. The worksheet was prepared using Wordsmith Tools (Scott, 1996) to find the citations in each article, so that students can be asked to compare citational practice across comparable texts in a narrow focus disciplinary context. In this instance we used a simple "catch-all" search string *19??)/??)*, that is, search for any five character string beginning with *19* -- remember this is a pre-21st century corpus -- and ending with a closing bracket, *and* any three character string ending with a closing bracket to catch other forms. Using this method, we located 112 citations in the 22 texts.

Table 8. Citation Worksheet

Example	Text	Cit. Cat
1. iety of possible surface realisations of that type of isomorphic relation we know as textual metaphor. <u>Christine Brooke-Rose (1958)</u> , in her A grammar of metaphor, provides a classification of forms of metaphor. However, her categorisation is unprinci	j7f	Integral Verb Controlling
2. information on the author from an asserted privileged position (much of the methodology and work of <u>F.R. Leavis (e.g. 1936, 1967)</u> and his followers is characterised by this approach). </p> <p> The two (realistic) perspectives remaining are positi	j7f	Non-Integral Refer
3. hur C. Clarke's 2001: a space odyssey, a novelisation of the film screenplay written by <u>Stanley Kubrick and Arthur C. Clarke (1968)</u> . </p> <p> This passage cannot be characterised as prototypical SF. It does not deal with aliens, space, technology,	j7f	Integral Naming
4. ism and claims to objectivity that have been increasingly questioned in the past twenty or so years (by writers from <u>Derrida (1975)</u> to <u>Lakoff (1987)</u> ). </p> <p> The middle way seeks to formalise, or at least make explicit, normative patterns in the	j7f	Integral Naming
5. of the overall communicative process involving an utterer and a receiver, very much in the implied spirit of <u>Grice's (1971, 1975)</u> projection of that communicative situation. However, the ordered pair of functions (f1 and f2) that are associated with	j7f	Integral Naming
6. intuitions. The value of the model is not only in recasting the traditional notion of the Co-operative Principle (from <u>Grice 1975</u> ), but also in describing the resolution of meaning as a principled negotiation between text and reader. The resolving st	j7f	Non-Integral Refer
7. though a holistic perspective is taken on literary stylistics in addressing science fiction. This approach follows <u>van Dijk (1977)</u> in regarding not only sentences but also textuality as the proper study of linguistics. In this, continental European	j7f	Integral Naming
8. ext-world to their cognitive universe (based on their previous familiarity with the patterns typical of the genre). <u>Eikmeyer (1989)</u> , in a paper from a conference on coherence, points out that reader interpretation depends on the depth of understandin	j7f	Integral Verb Controlling
9. cott, who filmed his reading of the novel as Blade runner in the early 1980s. Though <u>David Newnham, in the Guardian (24 July 1990)</u> , calls the film post- modernism: the movie, Scott's version is little more than a violent adventure story. Indiana Jone	j7f	Integral Verb Controlling
10. for example, involves judgements based on textual factors such as the narrative point of view ( <u>Fowler 1986: 127-46; Simpson 1990</u> ), the presentation of verbalisation ( <u>Leech and Short 1981: 318-51</u> ), the degree of non-actualised propositions ( <u>Leech 198</u>	j7f	Non-Integral Source
11. as often in literary discourse, author and reader are considerably separated by space and time. </p> <p> <u>Eikmeyer (1989: 27)</u> concludes his paper by introducing a subjectivity condition by which the participants judge the values of each parameter.	j7f	Integral Verb Controlling
12. the point of view of the reader's judgement of parameters is taken. Deriving from (and slightly correcting) <u>Eikmeyer (1989: 27)</u> , the prototypically co-operative parameters for the reader are: where J is the judgement or subjectivity condition. This	j7f	Integral Naming

The task for each group in Stage 2 is, therefore, to categorise the citations identified in each article, and then to pool results and present a summary of the range, purpose and forms of the citations that occur in this micro-corpus (the categorisations have been provided in this worked-up example).

In Stage 3 students will review their own citational practices (either using Wordsmith Tools to extract examples for analysis, or being provided with materials prepared by their teacher). Stage 4 will be on going and will involve a cycle of check-list supported peer review and self evaluation, supported by tutor comment on writing assignments or departmental work.

## CONCLUSION

In this paper we have described a range of citational practices in academic writing along with their linguistic realisations. We have also reviewed the extent to which published teaching materials provide learners with opportunities to develop their understanding of, and capacity to form, appropriate citations in their own writing, and found that, at the moment, these offer relatively little constructive support to apprentice writers. The need for such support has been underscored by a survey of a small number of EAP texts written by students on a pre-session course at a UK university. If teachers of English for Academic Purposes are to be able to help learners develop a better control of this essential academic writing knowledge/skill, we would recommend the accumulation of relevant collections of field specific texts as a resource for teachers and students of academic writing. By analysing these texts with word processing software or dedicated corpus tools (or by working with the results of teacher led analysis), students will be able to develop a fuller understanding of the cultural and linguistic role of citation in their fields of study and be much better placed to write well formed and appropriate academic texts.

## NOTES

1. Readers who wish to explore this area further may wish to start off with Aston (1996) or Tribble & Jones (1997).
2. We are grateful to one of our anonymous reviewers for the report that their undergraduate class had analysed the uses of citations in Hyland's (2000) book and found an above-average use of author-as-subject integral citations!
3. Potts is also the author of the article that this example is drawn from. In other words, this is a self-citation.
4. The categories presented here are a reduced set. The categories of Example (non-integral) and the three types of Verb-controlling (integral) citations, Research/Discourse/Other, in Thompson (2000) have been removed to make the explanation clearer.
5. Pre-session assignments written by postgraduate students on the following themes: EFL in Korea / Testing in ELT in Pakistan / Project implementation / Food industry / Agroforestry / International management
6. E.g., Caere corporation's **Omnipage Pro**® or ABBYY's Fine Reader®
7. E.g., WordSmith Tools (Scott, 1996) or MonoConc Pro (Barlow, 1999)
8. A text resource of 100 million words of late C20 British English that is now available internationally (contact <http://info.ox.ac.uk/bnc/> for more information).
9. The BNC file identifiers for the texts selected are J7F / J7G / J7H / J7J / J7K / J7L / J7M / J7R / J7S / J7T / J7U / J7V / J7W / J7X / J7Y / J80 / J81 / J82 / J83 / J84 / J85 / J86 / J87 / J88 / J89.

## ABOUT THE AUTHORS

Paul Thompson is a Research Fellow at the School of Linguistics and Applied Language Studies, Reading. He is currently studying towards a PhD, examining the language and organization of PhD theses in different disciplines.

e-mail: [p.a.thompson@reading.ac.uk](mailto:p.a.thompson@reading.ac.uk)

Chris Tribble is the author of *Writing* in the OUP teacher education series and has a long-term interest in the use of computers in text analysis and language description, written communication, and evaluation in education. He lives in Sri Lanka and lectures on the MA at King's College, London University.

email: [ctribble@sri.lanka.net](mailto:ctribble@sri.lanka.net)

## REFERENCES

- Aston, G. (1996). Corpora in language pedagogy: matching theory and practice. In G. Cook & B. Seidlhofer (Eds.), *Principle and practice in applied linguistics: Studies in honour of HG Widdowson* (pp. 257-270). Oxford, UK: Oxford University Press
- Barlow, M. (1999). *Monoconc Pro* [computer software]. Houston TX: Athelstan
- Bazerman, C. (1994). *Constructing experience*. Carbondale: Southern Illinois University Press.
- Becher, A. (1989). *Academic tribes and territories*. Milton Keynes, UK: Open University Press.
- Borg, E. (2000). Citation practices in academic writing. In P. Thompson (Ed.), *Patterns and perspectives: Insights for EAP writing practice* (pp. 14-25). Reading, UK: CALS, The University of Reading.
- Britton, J., Burgess, T., Martin, N., McLeod, A., & Rosen, H. (1975). *The development of writing abilities, 11-18*. London: Macmillan.
- Burnard, L., & McEnery, T. (Eds.). (2000). *Rethinking language pedagogy from a corpus perspective: Papers from the third international conference on teaching and language corpora (Lodz Studies in Language)*. Hamburg, Germany: Peter Lang.
- Campbell, C. (1990). Writing with others' words: Using background reading text in academic compositions. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom*. Cambridge, UK: Cambridge University Press.
- Connor, U. (1996). *Contrastive rhetoric: Cross-cultural aspects of second language writing*. Cambridge, UK: Cambridge University Press.
- Fox, H. (1994). *Listening to the world: Cultural issues in academic writing*. Urbana, IL: National Council of Teachers of English.
- Granger, S. (Ed.). (1998). *Learner language on computer*. Harlow, UK: Longman.
- Groom, N. (2000). Attribution and averral revisited: Three perspectives on manifest intertextuality in academic writing. In P. Thompson (Ed.), *Patterns and perspectives: Insights for EAP writing practice* (pp. 15-26). Reading, UK: CALS, The University of Reading.
- Hanania, E., & Akhtar, K. (1985). Verb form and rhetorical function in science writing: A study of MS theses in Biology, Chemistry and Physics. *ESP Journal*, 4(1), 45-58.
- Hill, S., Soppelsa, B., & West, G. (1982). Teaching ESL students to read and write experimental research papers. *TESOL Quarterly* 16, 333-347.

- Hyland, K. (1999). Academic attribution: Citation and the construction of disciplinary knowledge. *Applied Linguistics*, 20(3), 341-367.
- Hyland, K. (2000). *Disciplinary discourses: Social interactions in academic writing*. Harlow, UK: Longman.
- Johns, A. (1997). *Text, role and context*. Cambridge, UK: Cambridge University Press.
- Jordan, R. R. (1992). *Academic writing course*. London: Nelson.
- Malcolm, L. (1987). What rules govern tense usage in scientific articles? *English for Specific Purposes* 6, 31-44.
- Pennycook, A. (1996). Borrowing others' words: Text, ownership, memory, and plagiarism. *TESOL Quarterly*, 30(2), 201-230.
- Pickard, V. (1995). Citing previous writers: what can we say instead of "say"? *Hongkong Papers in Linguistics and Language Teaching*, 18, 89-102.
- Scott, M. (1996). *WordSmith Tools*. Oxford, UK: Oxford University Press.
- Shaw, P. (1992). Reasons for the correlation of voice, tense, and sentence function in reporting verbs. *Applied Linguistics*, 13(3), 302-319.
- Swales, J. M. (1981). *Aspects of article introductions*. Birmingham, UK: Aston University Languages Study Unit.
- Swales, J. M. (1986). Citation analysis and discourse analysis. *Applied Linguistics*, 7(1), 39-56.
- Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge, UK: Cambridge University Press.
- Swales, J., & Feak, C. (1994). *Academic writing for graduate students*. Ann Arbor, MI: University of Michigan Press.
- Thompson, P. (2000). Citation practices in PhD theses. In L. Burnard & T. McEnery (Eds.), *Rethinking language pedagogy from a corpus perspective*. Frankfurt: Peter Lang.
- Tribble, C., & Jones, G. (1997). *Concordances in the classroom: A resource book for teachers*. Houston TX: Athelstan.
- Tribble, C. (2001). Corpora and corpus analysis: New windows on academic writing. In J. Flowerdew, (Ed.), *Academic discourse*. Harlow, UK: Addison Wesley Longman.
- Trzeciak, J., & McKay, S. (1994). *Study skills for academic writing*. Hemel Hempstead, UK: Phoenix ELT.
- Weissberg, R., & Buker, S. (1990). *Writing up research: Experimental research report writing for students of English*. Englewood Cliffs, NJ: Prentice Hall Regents.