Citation Practices in Academic Corpora: Implications for EAP Writing

Su-Jung Min
(Kongju National University)

Min, Su-Jung. (2004) Citation practices in academic corpora: Implications for EAP writing *English Language & Literature Teaching*, *10*(3), 113-126

Explicit reference to the work of other authors is an essential feature of most academic research writings Corpus analysis of academic text can reveal much about what writers actually do and why they do so Application of corpus tools in language education has been well documented by many scholars (Pedersen, 1995, Swales, 1990, Thompson, 2000). They demonstrate how computer technology can assist in the effective analysis of corpus based data For teaching purposes, this recent research provides insights in the areas of English for Academic Purposes (EAP) The need for such support is evident when students have to use appropriate citations in their writings Using Swales' (1990) division of citation forms into integral and non-integral and Thompson and Tribble's (2001) classification scheme, this paper codifies academic texts in a corpus The texts are academic research articles from different disciplines. The results lead into a comparison of the citation practices in different disciplines Finally, it is argued that the information obtained in this study is useful for EAP writing courses in EFL countries

[EAP/corpus tools/genre analysis/citation practices, 전공 영어/코퍼스 도구/장르 분석/인용 형태]

I. INTRODUCTION

The development of English for Specific Purpose (ESP) involves close interaction between Applied Linguistics and English Language Teaching (ELT) ESP has been influenced by developments and changes in Applied Linguistics and ELT and it also has influenced on ELT ESP has been given its impetus by work in the area of Applied Linguistics, that is, register analysis, genre analysis, and discourse analysis Based on theoretical work in Applied Linguistics it has drawn on, ESP has been very influential in showing how a communicative language curriculum could be turned into a functional-notional syllabus or task-based syllabus (Dudley-Evans & John, 1998, Lackstrom, Selinker & Trimble, 1973), Swales, 1988) And ESP has traditionally been divided into two areas English for Academic Purposes (EAP) and English for Occupational Purposes (EOP) Research has been particularly strong in the area of EAP, where there is a growing interest in genre analysis (Bhatia, 1993, Swales, 1990)

Recent studies based on theoretical work in Applied Linguistics (Burnard & McEnery, 2000, Swales, 1990, Thompson, 2000) have shown that broad disciplinary areas such as Engineering, Business, and Medicine share a general pattern of organization and discourse features, and that there is significant variation between genres A genre is defined as a relatively fixed text-type and "a conventionalized verbal form associated with a conventionalized purpose of occasion" (Johnstone, 2002, p 156) The skills, understandings, or activities through which these text-types are called 'genre knowledge' (Berkenkotter & Hucken, 1995)

A genre analysis began with Swales' work (1981, 1990) on the introduction to an academic article Though critics such as Berkenkotter and Huckin (1995) argue that genre studies see genres as linguistic abstractions, it is the advantage of genre analysis "to relate textual findings to features of the discourse community within which the genre is produced" (Dudley-Evans & Jones, 1998, p 91-92). Among many one fairly stable genre in academic discourse is the scientific research article (Berkenkotter and Huckin, 1995, Swales, 1990, Tracy, 1988) Knowledge of genre in each disciplinary area is significant in writing academic texts because it involves an awareness of the

expectations and discursive practices of the discourse community.

Recent work demonstrates how corpus tools can be applied to language education in the field of EAP (Pedersen, 1995) In the situation where the subject courses are taught in the learners' first language like Korea, Japan and many other countries, EAP courses have focused almost exclusively on reading The materials focus on key micro-skills related to the overall macro-skill of reading, and certain lexical and grammatical items relevant to the comprehension of undergraduate academic reading texts (Alderson & Scott, 1992) Besides concentrating on comprehension of reading texts, as Swales (1986) suggests, it is necessary to expand the scope of EAP courses to the teaching of writing to post graduates who need to write research articles or theses

Academic research writing is a sophisticated artifact that displays a careful balance of factual information and social interaction (Bazerman, 1988, Hyland, 1999) Explicit reference to the work of other authors is an essential feature of most academic research writing, helping writers to establish a persuasive framework for the acceptance of their arguments But there are clear demarcations in the structure of subject-area knowledge system The incidence and use of citation might therefore be expected differ according to the norms and conventions for research presentation by academic discourse communities of each subject area

The authors need to follow the citation convention of academic discourse communities which they belong to when they present their research arguments Unlike professionals, however, the novice writers are not used to adopting the citation practices of a genre in certain academic community, that is, academic discipline. A person coming to know and to be able to function in academic community could learn a lot about the genre by analyzing sample texts This paper reports work that has used corpora to research citation practices of academic research writing from different disciplines and shows how current corpus tools can be used to give learners in EAP writing classes opportunities to extend their understanding of this aspect of academic discourse

II. CORPUS-BASED STUDIES ON CITATION PRACTICES

Corpus-based studies on citation practices in academic research articles were pioneered by Bazerman (1984) and elaborated by Swales (1986, 1990) Through the investigation of the historical changes in scientific research articles, Bazerman (1984) reported that there is considerable discoursal evidence in citational statements for growing abstraction, the deepening integration of present work within the relevant literature, and the increasing foregrounding of research as opposed to researcher From linguistic perspective, Swales (1990) divided citation forms into two clear categories integral and non-integral Citations forms that are integrated into a sentence and play an explicit grammatical role within the sentence are called integral Non-integral citations are usually placed within parenthesis and play no explicit grammatical role in the sentence The use of one form rather than the other reflects whether the emphasis is on the reported author or the reported work (Hyland, 1999)

Swales (1990) also used another dichotomous classification according to the verb form reporting and non-reporting In reporting citations the writer employs a 'reporting verb' such as 'show', 'establish', 'claim', etc, while non-reporting citations don't use such verbs Swales found that the ratio between non-reporting and reporting in science research articles was only 40-60, whereas it is a 25-75 ratio among literary critics Following Swales' classification of citation forms, Pickard (1995) used concordance software to investigate the citation practices of expert writers in the field of applied linguistics She also identified the different grammatical forms of integral citations

Using more sized corpora, Hyland (1999) investigated citations in a computer corpus of 80 research articles from different disciplines Based on the contextual variability of citations, Hyland showed the ways in which academic citation practices contribute to the construction of disciplinary knowledge Thompson (2000) also examined citation practices in doctoral theses from different fields Hyland and Thompson both investigated variation in frequency and patterning of citation practice in disciplinary discourse and found that the density of citations is lower among the science and technology article while the social science and humanity articles show higher citation rates But the doctoral theses

have much lower density of citation in Thompson (2000) compared to the research articles in Hyland (1999) The explanation given by Thompson and Tribble (2001) was that the types of texts produced in these two genres have different lengths Articles usually average between 2,000 and 5,000, while in Thompson's study, the average length of theses was between 31,000 and 63,000 Besides the frequency difference of citation, there is considerable variation in citation types between the different disciplines: integral and non-integral While both the science and technology disciplines and social science disciplines prefer non-integral forms over the integral, the former showed the much higher incidence of non-integral forms

As Thompson and Tribble (2001) pointed out, however, the formal distinction between integral and non-integral citations proposed by Swales (1990) does not provide practical help to EAP writers They need the range of choices available to them Thompson (2000) extended Swales' (1990) division of citation forms and presented a classification scheme Integral citations are subcategorized into verb controlling, naming, and non-citation, while non-integral citations into source, identification, reference and origin

III. CLASSIFICATION SCHEME OF CITATION FORMS

Basically adopting Thompson's (2000) scheme, this study modifies his scheme and classifies non-integral citations into source, identification and reference. Though Thompson (2000) distinguished source and origin, in this study these two are combined into one type of citation, source Source citations, according to Thompson and Tribble (2001), attribute a proposition to a source, while origin citations indicate the originator of a concept of a product But both where the idea comes from and where a concept or a product originate can be integrated into sources from the viewpoint of citation forms The main patterns of integral and non-integral citation forms are illustrated with examples in Table 1 and Table 2

As Table 1 illustrates, integral citations show the name of the researcher in a sentence In verb controlling citations, the citation acts as the agent that controls a verb, in active or passive voice Naming citation is a noun phrase or

a part of noun phrase Naming citations function as signifying work rather than the researchers In non-citation, the reference to another researcher is given without a year reference It is used when the reference has already been mentioned in the text and the writer does not need to repeat it

TABLE 1
Integral Citation Forms

Verb Controlling	By contrast, Baker, Gibbs, and Holmstrom (1994) report that level 6 managers in the firm they studied earned about five times the amount earned by level 1 mangers (E1)[1]
	The modern economic case for vouchers and increased educational choice was made by Milton Friedman(1962) (E2)
Naming	The flatness of the military structure is a puzzle because Rosen's hierarchy theory(1982) predicts that the larger the organization, (E1)
Non-citation	According to Bourdieu, in order to be considered an example of 'legitimate discourse', an utterance must meet (AL2)

TABLE 2
Non-integral Citation Forms

Source	Little or no attention was paid to developing the skills of listening, speaking, or communication, oral or written (Gorsuch, 1998), (AL1)
Identification	Inhibition of human aortic smooth muscle cell activation$^{by\ PPAR\ \alpha}$ ligands has also been reported recently (Staels et al) (LS2)
Reference	With teachers and their words as starting point (see also Li, 1998), the study applied (AL1)

Non-integral citations are usually placed within parenthesis and play no explicit grammatical role in the sentence Source citations indicate where the

1) (EI) indicates the source journal of the data in appendix

idea or a concept comes from That is, by attributing a proposition to another author, source functions as providing evidence for a factive report of findings or an idea. Identification citation identifies the author of the study it refers to In reference citations, the writer refers the reader to another research and the reference citation is signalled by the directive 'see' or 'refer' as in the example in Table 2

IV. DATA AND METHOD

This study is based on the analysis of citation practices in a corpus of research articles The corpus contains 96 research articles, consisting of 16 papers from each of 6 journals in four disciplines published during the period 1990-2003 from four different disciplines. Applied Linguistics, Economics, Life Science and Mechanical Engineering The journals were nominated by professionals in each field and the articles were selected at random from issues during the period The articles were scanned to produce a computer corpus Abstracts and references were excluded An electronic concordance, WordSmith Tools (Scott, 1996) was used to find the citations in each article The primary use of WordSmith Tool is to generate concordances or listings of all the occurrences of any given word in a given text, with words shown in context In this study, the canonical citational form, a date in brackets, was searched But this didn't account for a citation form in which the name is given without a year reference It is most commonly used form when the reference has been supplied earlier in the text and the writer does not want to repeat it So a concordance was made for all the names in the references of the articles, third person pronouns, and generalized terms for agents such as 'the author' and 'the researchers'. However, references to the thinking or schools (for example, 'Prague school') where a specific person was not referred to were excluded

V. RESULTS

The quantitative results show clear disciplinary variation in the extent to

which writers refer to the work of others and in how they choose different patterns of citations. The figures in Table 3 indicate the number of citations by discipline As the figures show, writers in Applied Linguistics and Economics employed substantially more citations than Life Science and Mechanical Engineering This supports the results of previous research (Hyland, 1999, Thompson, 2000, Thompson & Tribble, 2001) where the writers in the humanities and social science rely more on the work of others than scientists and engineers, with Mechanical Engineering and Life Science below the average However, it is interesting to note that the frequency and density of citations for Life Science were much lower than those in other studies And the figures of Mechanical Engineering (22 5 and 7 0) are lower than any of the figures

TABLE 3
Number of Citations by Discipline

Discipline	Av per paper	Av per 1,000 word
Applied Linguistics	69 3	9 7
Economics	78 2	8 9
Life Science	55 6	8 7
Mechanical Engineering	22 5	7 0
Overall Averages	56 4	8 6

Table 4 shows the relative percentages of the two types of citation, integral and non-integral These figures show that there was less variation in the ways, with all displaying a preference for non-integral forms But the figures show that there is also considerable variation in citation practice between the different disciplines In Life Science, it is interesting to find that the journal styles require numerical forms, which indicates the number in references at the end of the text All research articles in *Molecular and Cellular Biology* showed the same citation forms This kind of citation forms has an effect of reducing the prominence of cited authors

(1) In addition to the action as phagocytic and antigen-presenting cells, macrophages release a variety of inflammatory mediators that modulate

the immune response and encounter the infectious process (1, 47) (LS2)

In case of Applied Linguistics and Economics, writers showed the tendency to use adjunct agent structures, which give greater emphasis to authors This form never occurred in science and engineering research papers

(2) According to Schumaker and Lomax (1996), latent variable are psychological constructs that are inferred from empirical measurements (AL1)

TABLE 4
Ratios of Integral to Non-integral Citations

Discipline	Integral	Non-integral
Applied Linguistics	43 2	56 8
Economics	30 8	69 2
Life Science	1 8	98 2
Mechanical Engineering	17 2	82 8
Overall Averages	23 2	76 8

Table 5 shows the percentage of citations in the form of direct quotations Direct quotation is common in Applied Linguistics and Economics, but did not occur in Life Science and Mechanical Engineering at all

TABLE 5
Percentages of Citations in the Form of Direct Quotation

Discipline	Direct Quotations
Applied Linguistics	11
Economics	7
Life Science	0
Mechanical Engineering	0

Table 6 shows the relative percentages of the types of citation forms in different disciplines As Table 6 shows, writers in Applied Linguistics and Economics use integral verb controlling and non-integral source types much more frequently, while writers in Life Science and Mechanical Engineering make far greater use of non-integral source type The life scientists' preference is quite clear Total ratio of integral citation forms is considerably rare in Life Science research articles

TABLE 6
Percentages of All Types of Citation Form

	Integral			Non-integral		
	Verb Cont	Naming	Non-Cit	Source	Ident	Reference
Applied Linguistics	23 1	15 4	4 7	37 3	13 3	6 2
Economics	16 2	10 6	4 0	44 9	15 8	8 5
Life Science	0 0	1 2	0 6	62 8	28 2	7 2
Mechanical Engineering	0 0	15 2	2 0	56 1	23 4	3 3

The quantitative results suggest that there are clear disciplinary differences in the citation practices in research articles The types of research work and the conventions of the discipline influence the citation forms the writes in the fields choose (Thompson & Tribble, 2001) Reference to prior research clearly plays a more visible role in Applied Linguistics and Economics Similar results have been found in academic research writings and doctoral theses (Hyland, 1999, Thompson, 2000) The articles in Applied Linguistics and Economics comprised three fifths of all the citations in the corpus Writers in these fields were more likely to use integral forms of citation and emphasize the role of the researchers than in Life Science and Mechanical Engineering On the other hand, disciplinary conventions in Mechanical Engineering and Life Science did not attribute a stance to cited authors, and de-emphasizing the role of the researchers was particularly evident in the field of Life Science These findings are a little different from the results in Hyland (1999) Considering that *Cell and*

Molecular Biology in Hyland's study is close to Life Science in this study, it is quite interesting that the citation practices in two disciplines showed quite different results The relatively high frequency and density of citations in *Cell and Molecular Biology* in Hyland's study (82 7, 15 5) contrast with considerably low incidence in Life Science (55 6, 8 7) of this study. A possible explanation is that the selected journals for Hyland's study are from the fields of plant and cell. But the journals chosen in this study are more about the fields of biological chemistry and molecular biology The sub-disciplinary difference might cause the different results

VI. DISCUSSION AND CONCLUSION

Corpus analysis of academic texts revealed much about what writers actually do and why they do so Appropriate reference to previous work is an essential feature in academic writings as a means of supporting current claims The disciplinary differences in citation practices examined in this study suggest why writers choose one form rather than another depends on the inquiry patterns and knowledge structures of their discourse community Of course, the disciplinary variations cannot be seen as entirely determined and researchers are not simply passive recipients of conventions and norms of the discourse community they belong to But the patterns of citation explored in this study provide support for the view that writing practices are deeply embedded in the disciplinary conventions The inclusion of explicit references to the work of others is an essential feature of academic research writing, helping writers to establish persuasive framework for the acceptance of their arguments The incidence and use of citation differ according to norms and conventions for research presentation by academic discourse communities of each subject area The authors need to follow the citation conventions of academic discourse communities which they belong to when they present their research arguments

The quantitative results showed that there are substantial differences in citation practices between disciplines. The approach to the data included empirical observation, classification, and assessment of citation practices In this process, measurement was carried out quantitatively, that is, in the form of

absolute and relative frequencies of citation forms What do this results mean for teaching EAP learners? The information obtained and described in this study can be assessed for EAP development In EAP we need to introduce the idea that different disciplines expect students to adopt different stance The models provided for EAP students have been focused in reading (Dudley-Evans & John, 1998) and in teaching of writing in EAP the emphasis has been on summary, paraphrase, and quotation (Thompson & Tribble, 2001) So, arguments for the development of corpus-informed approach in EAP have been made by Hyland (2000) and Thompson and Tribble (2001) With the help of corpus of professional academic writings in their field, novice writers can develop their understanding of more appropriate forms of citation in their writings From pedagogical point of view it is possible to convert the analysis into teaching material that provides learners with opportunities to develop their understanding of, and capacity to form, appropriate citations in their own writing If teachers of EAP are to be able to help learners develop a better control of essential academic writing skill, it is recommended that the accumulation of relevant collections of field specific texts as a resource for teachers and students of academic writing By analysing these texts with dedicated corpus tools, students will be able to develop a fuller understanding of the cultural and linguistic role of citation in their fields of study and be better placed to write well formed academic texts

This study has neglected a number of issues that provide ground for further research There are attribution practices in other disciplines and genres that are not examined in this study, and it is also probable that closer analyses will reveal variations in sub-genres and sub-disciplines Comparative analysis of citation forms of qualitative and quantitative research papers within same discipline will reveal interesting results

REFERENCES

Alderson, J, & Scott, M.. (1992). Insiders, outsiders and participatory evaluation. In J Alderson & A Beretta (Eds), *Evaluating second language education.* (pp 25-28) Cambridge University Press

Bazerman, C (1984). Modern evolution of the experimental report in physics Spectroscopic articles in Physical Review, 1893-1980 *Social Studies in Science 14*, 163-196

Bazerman, C (1988) *Shaping written knowledge* Madison: University of Wisconsin Press

Berkenkotter, C., & Huckin, T (1995) *Genre knowledge in disciplinary communication Cognition/culture/power* Hillsdale, NJ Lawrence Erlbaum

Bhatia, V (1993) *Analyzing genre* London: Longman

Burnard, L., & McEnery, T (Eds.) (2000) *Rethinking language pedagogy from a corpus perspective: Papers from the Third International Conference on Teaching and Language Corpora.* Hamburg, Germany Peter Lang

Dudly-Evans, T., & John, M (1998) *Development in English for specific purposes* Cambridge, UK Cambridge University Press

Hyland, K (1999) Academic attribution Citation and the construction of disciplinary knowledge, *Applied Linguistics, 20*(3), 341-367

Lackstrom, J E., Slinker, L, & Trimble, L. (1973) Technical rhetorical principles and grammatical choice *TESOL Quarterly, 7*, 127-136

Pedersen, J (1995) The identification and selection of collocations in technical dictionaries *Lexicographia, 11*, 60-73

Pickard, V (1995) Citing previous writers What can we say instead of "say"? *Honkong Papers in Linguistics and Language Teaching, 18*, 89-102

Scott, M (1996) *WordSmith* Oxford, UK Oxford University Press

Swales, J (1981) *Aspects of article introduction* Birmingham, UK Aston University Languages Study Unit

Swales, J (1986) Citation Analysis and discourse analysis *Applied Linguistics, 7*(1), 39-56

Swales, J (1988) *Episodes in ESP* Hemel Hempstead Prentice Hall International

Swales, J (1990) *Genre analysis English in academic and research settings* Cambridge, UK Cambridge University Press

Thompson, P (2000) Citation practices in Ph.D theses In L Burnard & T McEnery (Eds.), *Rethinking language pedagogy from a corpus perspective* (pp 91-101) Frankfurt Peter Lang

Thompson, P, & Tribble, C (2001) Looking at citations Using corpora in English for academic purposes, *Language Learning & Technology* 5(3), 91-105

APPENDIX
JOURNAL CORPUS

Applied Linguistics (AL)
1 TESOL Quarterly
2 Applied Linguistics
Economics (E)
1 Journal of Labor Economics
2 The American Economic Review
Life Science (LS)
1 Molecular and Cell Biology
2 Journal of Biological Chemistry
Mechanical Engineering (ME)
1 Journal of Heat Transfer
2 Journal of Fluid Mechanics

예시언어(Examples in): English
적용가능 언어(Applicable Language): English
적용가능 수준(Applicable Levels) College/Higher

Min, Su-Jung
공주대학교 사범대학 영어교육과
314-701 충남 공주시 신관동 182번지
Tel 041-850-8190
Fax 041-850-8190
Email sujmin@kongju ac kr

Revised in Oct, 2004
Reviewed by Nov, 2004
Revised version received in Dec, 2004