

A Corpus-Based Approach to Citation Use in English for Academic Purposes*

Myung-Jeong Ha
Sangmyung University

Ha, Myung-Jeong. (2012). A Corpus-based approach to citation use in English for academic purposes. *Modern English Education*, 13(2), 37-55.

Academic writers are expected to make explicit references to the previous literature that is an essential part of academic writing. Apprentice writers may undergo a troublesome experience because of the lack of awareness in citation use. Following Swales' (1990) distinction between integral and non-integral citation forms, this study examined citation patterns particularly focusing on reporting clauses. It investigated finite reporting clauses with *that*-clause complement. The texts are 15 research articles written by international writers in the field of applied linguistics. Using the WMatrix, the corpus was automatically tagged, by part-of-speech tags with CLAWS7. The corpus then was analyzed using AntConc tools that can offer a powerful concordancer. The results of analyses revealed that the academic research papers frequently used reporting clauses and that they generally occurred as a type of integral citations with a human subject. The findings of the study suggested how working with citation patterns can be of great use in raising students' language awareness. This study underlines the importance of effective use of citation that may contribute to students' academic success, suggesting the application of corpus tools in classroom activities from a pedagogical point of view.

[citation/academic writing/corpus/English for Academic Purposes/
인용/학술영작문/코퍼스/학술영어]

I. INTRODUCTION

Academic writing is a sophisticated artifact that presents a careful balance of factual information and social interaction (Hyland, 1999). One of the most significant current discussions in academic writing is citation use. It allows the writer to acknowledge with the

* This study was supported by intramural research funds from Sangmyung University in 2012.

contributions of other scholars and to establish one's own academic authority and credibility. Specifically the research writer's concern for audience is closely connected with reference to prior research, reminding us that statements are a response to previous statements and are available for further statements by others – that is, “the attribution of propositional content to another source” (Hyland, 1999, p. 341). The use of citation may vary according to the conventions for research presentation by discourse communities. The writers in a certain subject area are expected to follow the citation conventions shared by the community members. Also it is necessary for novice writers to become familiar with the citation practices of a genre in disciplinary writing.

Within applied linguistics, the literature on citation has explored cultural variation in citation use (Pennycook, 1996), disciplinary differences in citation use (Hyland, 2000), and reporting verbs accompanying citation (Hyland, 2002; Shaw, 1992). The wide range of the literature shows the complexity of citation practices and this helps to explicate plagiarism (Pecorari, 2003), difficulties in expressing one's voice, and unclear distinction between one's own points and those of other authors (Borg, 2000).

However, little attention has been paid to citation from a perspective of integral and non-integral distinction. Since the distinction between integral and non-integral citations is the most frequently used in terms of citation typology, the present study builds on the Swales' citation forms. Also Swales' citation forms generally employ corpus-linguistic methods combined with corpora. In addition, attention to the type distinction related to function may tell us about ways in which expert writers choose a particular citation type to their specific purposes. Therefore, the present paper shows a case of integral choices against non-integral citation that may inform apprentice writers of choosing relevant citation forms to their specific needs.

This paper focuses on the reporting clause as its starting point since this allows for identification of the frequent patterns used in citation and thus offers linguistic input to the concept of grammatical consciousness-raising (Hunston & Francis, 2000). The goal of this particular aspect of the study was to investigate L2 writers' citation practices regarding the use of language taken from published papers. Specifically, I intend to analyze the use of reporting clause (i.e., the verb patterns *V that*) in research articles.

In the following sections, I review relevant literature and present types of citation forms of which part is primary analytical framework of the study. I report work that has used corpora to investigate a particular aspect of academic writing – that is, citation use. I then describe the corpus data and the analytic procedures. The next section presents the major findings of the study, before concluding the paper with a discussion on pedagogical implications for EAP (English for Academic Purposes) instructors and students and some directions for future research.

II. THEORETICAL BACKGROUND

1. Corpus-Based Approaches to Citation Practices

In this paper, a citation is viewed as “a research report which has a specific reference point that is clearly identifiable” (Charles, 2006, p. 314). Various typologies have been developed due to many researchers’ attempts to classify citations.

Swales (1990), for example, conducted the study of citation analysis. He distinguished integral citation forms from non-integral citation forms. Integral citation is placed within the sentence and plays an explicit grammatical role while non-integral citation is outside the sentence, typically placed within the brackets, and plays no explicit role within the structure of the sentence. The use of one form rather than the other represents a choice to give greater stress on either the reported author or the reported text.

Alongside Swale’s work, there has been substantial research into genres and disciplines (e.g., Hyland, 1999; S. J. Min, 2004; E. Shim, 2006; Thompson, 2000). Hyland (1999), for instance, investigated citations in a corpus of 80 research articles that were composed of 10 journal articles from different disciplines. He reported variation in practice in disciplinary discourses and made use of frequency and concordance data to present patterning across the texts. The lower density of citations among the science articles contrasted with the higher one among the social science articles. In the same vein, Thompson (2000) studied citation forms in PhD theses from different academic areas. The study revealed distinct differences in the use of citations in theses in two subdisciplines such as Agricultural Botany and Agricultural and Food Economics. It was reported that writers in Agricultural Botany used the non-integral source much more frequently, whereas the Agricultural Economics had a preponderance of integral citations. Drawing upon Thompson’s scheme, S. J. Min (2004) investigated citation preferences in 4 different disciplines (i.e., Applied Linguistics, Economics, Lite Science, and Mechanical Engineering). She claimed that there are clear disciplinary differences in the citation use in research articles.

Research on citation also dealt with verb tense and voice on the reported information (e.g., Myers, 1992; Shaw, 1992; Swales, 1990). For example, Swales (1990) stated that the present simple-present perfect-past scale, covering more than 90 % of finite reporting verbs, shows increasing distance of various kinds from the reported finding. Another important rhetorical choice has been the use of a reporting verb that introduces the work of other researchers (Hunston, 1993; Thompson & Ye, 1991). The function of those reporting verbs is to allow the writer to convey the activity reported and to distinguish an attitude to that information.

When it comes to citation in L2 student writing, researchers have mainly focused on the issues of paraphrasing and summarizing (Campbell, 1990), plagiarism (Howard, 1995),

and unclear distinction between one's own points and those of other authors (Groom, 2000). To summarize, the wide range of the citation literature reflects the complex nature of citation practices, indicating that novice writers may experience difficulties in learning how to cite. One criticism of much of the literature on citation is that the existing accounts fail to look at citations from a pattern perspective. The approach of researchers has been to count all occurrences of citation and to analyze them in terms of their structure and function. By contrast, this paper attempts to show the verb patterns *V that* and *it be V-ed that*.

2. Types of Citation Forms

Due to different aspects of citation use, citation typologies can be divided into content-based typologies and form-based typologies. A researcher needs the knowledge of the subject matter and double-rating procedures to develop content-based typologies. Moravcsik and Murugesan (1975), for example, proposed four parameters of contrasting features of citations (see Table 1). The consequent uses of this typology have shown that some of the categories are irrelevant to other disciplines such as applied linguistics.

TABLE 1
Design of the Experiment

Category	Function
Conceptual (or) Operational	Is the reference made in connection with a concept or theory that is used in the referring paper?
Organic (or) Perfunctory	Is the reference truly needed for the understanding of the referring paper or to the working out of the content of that paper?
Evolutionary (or) Juxtapositional	Is the referring paper built on the foundations provided by the reference (or) Is it an alternative to it?
Confirmative (or) Negational	Is it claimed by the referring paper that the reference is correct (or) Is its correctness disputed?

On the other hand, citation typologies based on formal criteria focused on the surface forms of citations rather than the meanings. A representative example is the Swales' (1990) typology, which consists of integral and non-integral citations. In integral citations, the name of the cited author is placed within the sentence, whereas, in non-integral citations, the author's name is placed either in parentheses or in footnotes/endnotes. The following examples (Swales, 1990, p. 149) demonstrate the two forms:

(1) Integral citation:

Brie (1988) showed that the moon is made of cheese.

Brie's theory (1988) claims that the moon is made of cheese.

(2) Non-integral citation:

Previous research has shown that the moon is made of cheese (Brie, 1988).

It has been shown that the moon is made of cheese (Brie, 1988).

Further, Thompson (2001) divided integral and non-integral citations into sub-types. His classification considered both formal linguistic criteria and a functional aspect (see Thompson, 2005 for further details). Based on the formal criteria, integral citations are thus divided into: (a) verb controlling (if the citation controls a lexical verb); (b) naming (where the citation is a noun phrase or its part); and non-citations (where the author's name is not followed by other data). Non-integral citations are divided into: (a) source, which indicates where the information is taken from; (b) identification citations, which identify an actor in the sentence; (c) origin citations indicating the originator of a concept or product; (d) reference citations, which point to the work containing further information; and (e) example, where the cited work shows what is stated in the sentence.

Based on Thompson's (2001) classification, Petric's (2007) citation classification included attribution, exemplification, further reference, statement of use, application, evaluation, establishing links between sources and comparison of one's own findings or interpretation with other sources. Some of the citation functions state whether they are embedded in integral or non-integral citation. According to Petric, for instance, "both integral and non-integral citations can have this rhetorical function" (p. 243). With respect to potential contextual differences in integral and non-integral citations, it is well known that non-integral citations de-emphasize the role of the reported author and put emphasis on the reported content.

Up to now, EAP course books have not provided enough information about citation use (Thompson & Tribble, 2001). In the same line, Petric (2007) stated that "students have difficulty using sources in academically acceptable ways" (p. 239). This may be the case with Korean graduate students in English academic writing

III. METHOD

1. Corpus

This study is based on a corpus of 15 research articles, consisting of one paper from each of three leading journals in applied linguistics (see Appendix): TESOL quarterly, Applied Linguistics, and English for Specific Purposes. The articles were written by international

English writers looking at English as a lingua franca and were chosen from issues published in the years of 2010 and 2011. By focusing on the papers from applied linguistics, the researcher attempted to extend the view that different disciplines can affect the manner of expression in developing arguments (Hyland, 2004). Although whole texts were used, they excluded all tables, figures, and bibliography. Using the UAM Corpus Tool¹ (O'Donnell, 2008), general statistics of the corpus was produced. After preparation of the text, the word counts for the fifteen articles in the study are shown in Table 2.

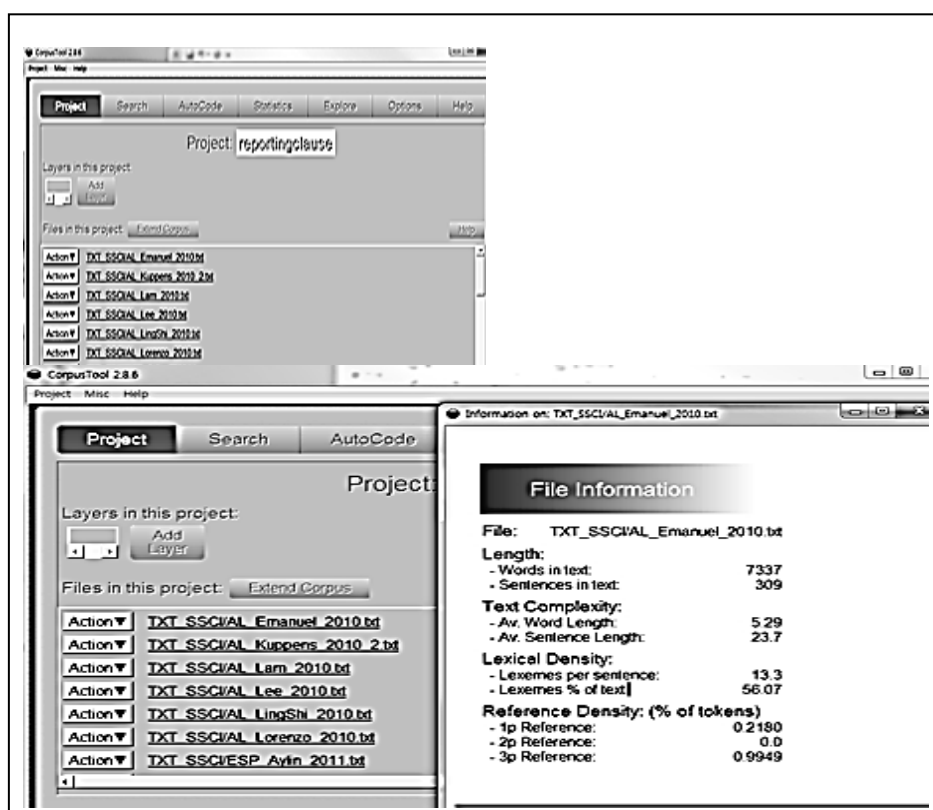


FIGURE 1 The Screenshot of the UAM Corpus Tools

¹ <http://www.wagsoft.com/CorpusTool/index.html>.

TABLE 2
General Statistics of the Corpus

Text	Length Words in Text	Text Complexity Ave. Sentence Length	Lexical Density Lexemes per Sentence
1	7337	23.7	13.3
2	7243	21.8	12.5
3	7368	17.8	10.0
4	6222	19.5	1.2
5	7056	17.4	9.70
6	6193	20.3	12.0
7	6380	12.6	7.27
8	6511	21.7	12.7
9	6190	20.7	12.2
10	7423	11.2	6.45
11	6208	22.9	13.6
12	7685	17.3	9.31
13	7149	17.0	9.54
14	7304	22.3	12.8
15	7385	17.5	10.8

Note: A lexeme means a set of forms taken by a single word. For example, suppose, supposes, supposed, and supposing are forms of the same lexeme, SUPPOSE.

2. Procedure

In the initial steps, the corpus was tagged using the CLAWS tagger in WMatrix² (Rayson, 2001) to facilitate recovery of instances of *that* complementizer. CLAWS has consistently attained 95-98% accuracy (Garside & Smith, 1997). Using the Wmatrix the corpus was automatically tagged, by part-of-speech tags with CLAWS³. Using AntConc Tools (Anthony, 2004), concordances were created on the search word *that* and finite reporting clauses were identified. Following the prior research that showed the retention of *that* complementizer as the norm in academic writing (Biber, Johansson, Leech, Conrad & Finegan, 1999), concordances were made on the search word *that* and finite reporting

² <http://ucrel.lancs.ac.uk/wmatrix/folders.png>.

³ Refer to Garside and Smith (1997).

clauses were identified (see Figures 2 and 3).

Next, clauses that report on the studies of other researchers were distinguished. Using terminology drawn from Charles (2006), these *research reports* are seen as reports with a source which is a reference to the research of others in applied linguistics. The research reports are identified by (a) the presence of a grammatical subject which is a (pro)noun reference to a researcher or researchers (e.g., Fung and Carter; they); (b) alternatively either a textual feature or the surrounding context indicates that the proposition is taken from others' research (e.g., an increasing number of studies; his study; it is suggested that ... ; it has been increasingly acknowledged that ...).

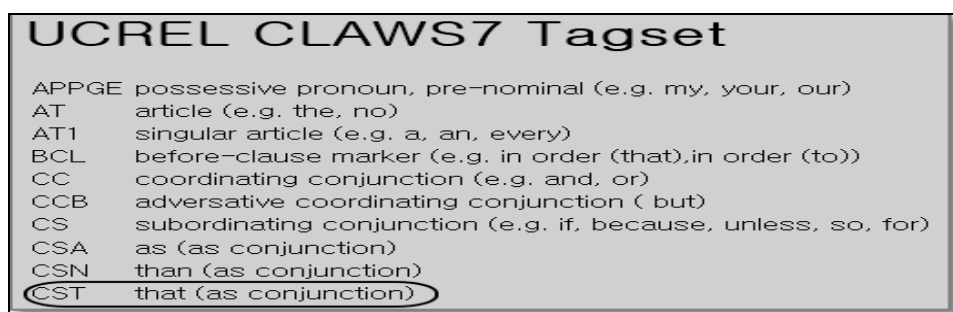


FIGURE 2 The Screenshot of Tagset in WMatrix

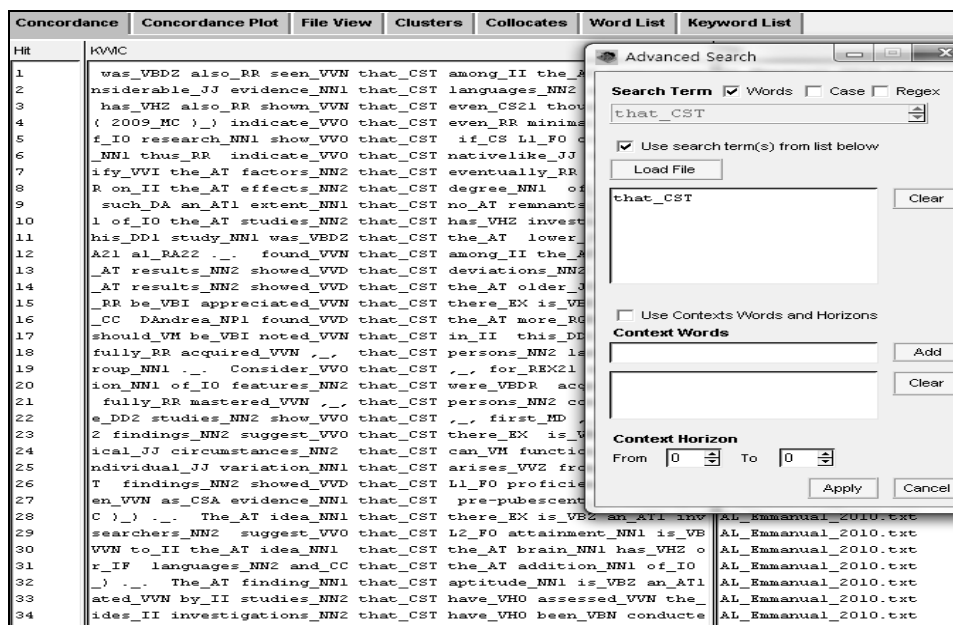


FIGURE 3 The Screenshot of the Concordances of That as Conjunction

Third, two types of further analysis are performed: First, I distinguished three clause types according to the grammatical subject of the reporting clause: a noun group with human reference; a noun group with non-human reference; and introductory it followed by passive voice:

- (i) *Fung and Carter (2007) report that native speakers use a wider range of discourse markers including actually, right, and well to achieve a broader variety of pragmatic functions such as marking responses and attitudes.*
- (ii) *Research has also shown that even though exposure to a given language before puberty is a crucial condition, ...*
- (iii) *It has been suggested that an additional processing load is induced for subject extraction by backtracking (Schachter & Yip 1990) ...*

Fourth, following Swales (1990), non-integral and integral citation forms are identified. As mentioned previously, non-integral forms are citations that are outside the sentence and which play little explicit grammatical role in the sentence, while integral forms are those that play an explicit grammatical role within a sentence. In this study, however, I identify only citations that use a reporting clause with *that*-complement, relying on the categorization of follow-up sentences according to the form of the original citation⁴. Integral citation in the reporting clauses is illustrated below:

- (iv) *Hakuta and D'Andrea found that the more frequently the Spanish was spoken at home, the higher the participants' scores were on the language tests.*
- (v) *The studies by DeKeyser (2000) and Abrahamsson and Hyltenstam (2008) provide information about ...*
- (vi) *Findings from e.g. Ruben (1999) and Hyltenstam et al. (2009) indicate that even ...*

Non-integral citations place the author's name outside of the text, either in parentheses or in footnotes. Non-integral citation in each of the three clause types is demonstrated below:

⁴ For more details, refer to Shaw (1992).

- (vii) *A number of researchers found that students relied on source texts in their writing (Campbell 1990; Moore 1997; Shi 2004).*
- (viii) *An increasing number of studies have pointed out that the use of discourse particles contributes to the pragmatic and communicative competence of speakers (Wierzbicka 1991; Müller 2005).*
- (ix) *It has been increasingly acknowledged that discourse particles are crucial for learners to communicate ...*

Also the type of reference involved was examined, resulting in two categories of research reports such as “citation” and “general reference” are identified. Here general reference is defined as a category of research reports without a specific reference point.

IV. RESULTS

1. Three Clause Types

Table 3 below presents the distribution of the finite reporting clauses with *that*-clause complement throughout the fifteen articles. As can be seen in Table 3, texts 3 and 11 show relatively lower figures of finite reporting clauses with *that*-clause (5 and 6 respectively). It appeared that the two texts commonly contained more noun phrases followed by *that* clause (e.g., the observation that, the fact that, and the image that, etc).

TABLE 3
The Distribution of Finite Reporting Clauses with *That*-clause Complement
Used in 15 Articles

Text	Number of Words in Text	Number of Finite Reporting Clauses with <i>That</i> -Clause
1	7337	31
2	7243	12
3	7368	5
4	6222	16
5	7056	12
6	6193	11
7	6380	26
8	6511	8
9	6190	11
10	7423	8
11	6208	6

12	7685	11
13	7149	8
14	7304	12
15	7385	10

Table 4 presents the frequency of three clause types in each article. It shows that overall human subjects are the most frequent, followed by lower numbers of non-human subjects. *It* subject records the lowest frequency figures. The construction of knowledge may vary according to the academic culture of discourse communities. The result indicates that the discourse in the Humanities and Social Sciences including applied linguistics tends to be more personal and subjective, as previous studies (Becher & Trowler, 2001; Charles, 2006) revealed.

TABLE 4
Frequency of Three Clause Types

Text	Human Subject	Non-Human Subject	<i>It</i> subject with passive
1	16	12	3
2	8	4	0
3	3	1	1
4	8	6	2
5	8	4	0
6	7	2	2
7	22	2	2
8	11	5	0
9	13	3	0
10	5	2	1
11	5	1	0
12	8	3	0
13	6	2	0
14	8	3	1
15	7	3	0

It was found that there is a more or less strong tendency to use reporting clauses to make integral citations with a human subject. It is suggested that the frequent use of integral citation with a human subject results in prominence of the cited researcher (Charles, 2006). This finding, however, is not consistent with the results of the previous research (e.g., Hyland, 2002). Hyland (2002) found that non-integral citation was predominant in the research articles he examined. By contrast, Thomson's (2001) study showed somewhat mixed results in that integral citation was salient in agricultural economics while non-

integral citation was preferred in agricultural botany. It appears that the corpus data in the present study refers only to reporting clauses, whereas Hyland and Thompson are concerned with all types of citations. Accordingly, *V that* pattern tends to be used when constructing an integral citation with a human subject. Calculations for the data under this study demonstrate that, in each article, more than 62% of the integral citations examined use a human subject (see Table 5).

TABLE 5
Frequency of Reference and Clause Types

Text	Integral citation			Non-integral citation			General reference		
	Human subject	Non-human subject	<i>It</i> subject with passive	Human subject	Non-human subject	<i>It</i> subject with passive	Human subject	Non-human subject	<i>It</i> subject with passive
1	16(62%)	7(27%)	3(12%)	0	4	0	0	1	0
2	5(83%)	1(17%)	0(0%)	3	2	0	0	1	0
3	2(67%)	1(33%)	0(0%)	1	0	1	0	0	0
4	7(78%)	1(11%)	1(0%)	1	3	1	0	2	0
5	6(100%)	0(0%)	0(0%)	2	4	0	0	0	0
6	5(100%)	0(0%)	0(0%)	2	0	2	0	2	0
7	22(96%)	1(4%)	0(0%)	0	1	2	0	0	0
8	11(79%)	3(21%)	0(0%)	0	1	0	0	1	0
9	11(85%)	2(15%)	0(0%)	2	1	0	0	0	0
10	3(75%)	1(25%)	0(0%)	2	1	1	0	0	0
11	5(83%)	1(17%)	0(0%)	0	0	0	0	0	0
12	8(100%)	0(0%)	0(0%)	0	2	0	0	1	0
13	4(67%)	2(33%)	0(0%)	0	0	0	2	0	0
14	5(83%)	1(17%)	0(0%)	3	1	0	0	1	0
15	6(86%)	1(14%)	0(0%)	1	1	0	0	1	0

Note: bracketed number = frequency percentage (calculated against the total number of integral citations)

As a reporting verb with *that*-complement is most frequently used in integral citation with a human subject, this paper highlights this pattern. Generally this type of citation is supposed to be connected with comment on a particular study rather than with generalizations about many studies (Charles, 2006). As Charles (2006) stated, writers hardly subsume many papers under a single sentence. Rather, they have a tendency to deal with the literature by discussing certain papers at some length. Writers may show full

mastery of the literature by discussing specific studies in detail and, at the same time, situate cited researchers within the field. This is demonstrated as follows:

(x) *Young and Gaea (1998) suggest that citation skills develop along with one's disciplinary knowledge and students achieve knowledge transformation by integrating content as interpreted evidence. ... Campbell (1990) noted that copying was a major strategy used by both L1 and L2 writers. ... Based on comparisons of passages from student texts and the cited sources, Pecorari (2003) found that students' citations were not transparent ...*

However, these studies have either provided a limited description of students' text in terms of quantities and weightings of various types of use of citations or textual appropriation, ... What is missing here is a great amount of textual borrowing not indicated by citations.

To tap these issues, the present study uses undergraduates' self-reflections ... (AL5)

The first three sentences display some of the extended discussions of the literature by using three integral citations with human subjects. A further paragraph that details problems in the previous literature comes along. In the last sentence, finally, the writer adopts the eclectic position to situate his or her research as an attempt to fill the created gap in the current knowledge.

V. CONCLUSION

While previous research has mainly employed L1 corpora, the present study attempted to examine the patterns of citations in an L2 corpus of applied linguistics articles. The study explored reporting clauses with a *that*-clause complement in reporting others' research. The most frequently observed pattern was an integral citation with a human subject. While Hyland (2002) found a high proportion of non-integral citations, the corpus I examined showed higher figures for integral citations. Whereas the results of Thompson (2001) were correspondent with those of Hyland for theses in agricultural botany, the greater prominence in agricultural economics was integral citation. As discussed in the previous section, the variation of these findings seems to be explained by the fact that the data used in the study exclusively focus on reporting clauses, whereas all citations are considered in both Thompson's and Hyland's studies.

The integral citation use with a human subject is generally used by the writers in order to place the others' work within the field so that they can position themselves in relation to it. According to Charles (2006), certain features make this pattern well-tuned to situating

writers themselves and others: (a) the use of reporting clause that opens up a space allowing comment on the cited proposition; (b) the use of integral citation with a human subject facilitating comment on individual works. The result of the study is also consistent with Thompson and Tribble (2001), supporting that thesis writers generally have a preference to deal with the literature review by discussing individual papers rather than by subsuming several papers within a single comment.

This study has the following pedagogical implications. First of all, it is important to recognize the value of citation use particularly in EAP courses. When it comes to language re-use in writing for publication in English, researchers have mainly highlighted its troublesome matters such as plagiarism (Howard, 1995) and difficulties in expressing one's voice (Borg, 2000). Although students have had difficulties making references to the literature in academically acceptable ways, there has been lack of explicit guidance on what constitutes legitimate citation practices in academic writing. Attention to citation patterns is of great use in that it reveals recurrent forms of expression. Specifically, it is important for students to recognize that one method of incorporating citation into text is to use a reporting verb followed by a *that*-clause with a human subject. In addition, effective use of citation may help stress the knowledge valued by students and thus contribute to students' academic success. It is suggested that EAP instructors can develop students' citation skills by raising their awareness of the rhetorical functions that is completed through citation. One possible approach with regard to classroom activities is to encourage students to discuss the writer's intentions behind citation use from different research articles.

Secondly, with the help of corpus tools in English language education, EAP instructors can produce concordanced exercises that focus on citation patterns used within the students' discipline areas. Concordancers have been shown to be an effective aid in facilitating the learning of vocabulary, collocations, grammar, and writing styles (Sun & Wang, 2003). For example, AntConc is a simple and easy to use corpus analysis toolkit that has been shown to be effective in a classroom context (Anthony, 2004). As Anthony suggested, AntConc seems to be designed specifically for students in a classroom context. As part of a data driven approach to learning (Johns, 1997), students themselves can search for citation patterns which are otherwise not apparent. This approach is highly recommended if students can analyze their course books because of their familiarity with the course content. By doing this, students will be able to have a fuller understanding of the role of citation in their areas of study and thus produce academically appropriate texts.

For the future direction of research in citation, it seems relevant to resort to Flowerdew and Li's (2006) stark observation: First, to trace a longitudinal dimension apprentice writers' development in citation use over time and to see what changes occur as they become enculturated into their target discourse community as competent writers. Second,

to investigate the citation practices of scholars across disciplines in L2 contexts.

More extensive investigation into the role of citation is thus warranted in a wider range of discourse communities that are subject to a more detailed account of the phraseology of citation. Also it is necessary to compare citation use by competent writers and by apprentice writers in the same discourse community in order to reveal citation functions that are appropriate to student writing.

Finally, the limitation of the study has to be mentioned. Given that the corpus data of the present study consists of 15 research articles, there may be a limitation to generalization of its findings. Nevertheless, the prominent significance of the study is to gain insight into the extent and type of patterning found in academic research articles. Considering that there have been few scholarly discussions on corpus-based citation patterns in terms of L2 academic writing, it is hoped that the study reported in this paper has been a useful contribution for EAP practitioners, or more broadly for TESOL professionals in the Korean EFL context.

REFERENCES

- Anthony, L. (2004). AntConc: A learner and classroom friendly, multi-platform corpus analysis toolkit. *Proceedings of IWLeL 2004: An interactive workshop on language e-learning, Japan*, 7-13.
- Becher, T., & Trowler, P. (2001). *Academic tribes and territories: Intellectual enquiry and the cultures of disciplines* (2nd ed.), Buckingham: SRHE and Open University Press.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow: Pearson Education.
- Borg, E. (2000). Citation practices in academic writing. In P. Thompson (Ed.), *Patterns and perspectives: Insights into EAP writing practices* (pp. 27–45). Reading: University of Reading.
- 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* (pp. 211–230). Cambridge, UK: Cambridge University Press.
- Charles, M. (2006). Phraseological patterns in reporting clauses in citation: A corpus-based study of theses in two disciplines. *English for Specific Purposes*, 25(4), 310–331.
- Flowerdew, J., & Li, Y.-Y. (2006). Language re-use among Chinese apprentice scientists writing for publication. *Applied Linguistics*, 28(3), 440-465.
- Garside, R., & Smith, N. (1997). A hybrid grammatical tagger: Claws 4. In R. Garside, G.

- Leech & A. McEnery (Eds.), *Corpus annotation: Linguistic information from computer text corpora* (pp. 102–121). London: 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: University of Reading.
- Howard, R. M. (1995). Plagiarism, authorships, and the academic death penalty. *College English*, 57(7), 788-806.
- Hunston, S. (1993). Evaluation and ideology in scientific writing. In M. Ghadessy (Ed.), *Register analysis: Theory and practice* (pp. 57–73). London: Pinter.
- Hunston, S., & Francis, G. (2000). *Pattern grammar: A corpus-driven approach to the lexical grammar of English*. Amsterdam: Benjamins.
- 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*. Michigan: The University of Michigan Press.
- Hyland, K. (2002). Activity and evaluation: Reporting practices in academic writing. In J. Flowerdew (Ed.), *Academic discourse* (pp. 115–130). Longman: Harlow.
- Hyland, K. (2004). Disciplinary interactions: Metadiscourse in L2 postgraduate writing. *Journal of Second Language Writing*, 13, 133–151.
- Johns, T. (1997). Contexts: The background, development and trialling of a concordance-based CALL program. In A. Wichmann, S. Fligelstone, T. McEnery & G. Knowles (Eds.), *Teaching and language corpora* (pp. 100-115). London: Longman.
- Min, Su-Jung. (2004). Citation practices in academic corpora: Implications for EAP writing. *English Language & Literature Teaching*, 10, 113-126.
- Moravcsik, M. J., & Murugesan, P. (1975). Some results on the function and quality of citations. *Social Studies of Science*, 5, 86–92.
- Myers, G. (1992). ‘In this paper we report ...’: Speech acts and scientific facts. *Journal of Pragmatics*, 17, 295-313.
- O'Donnell, M. (2008). The UAM Corpus tool: Software for corpus annotation and exploration. *Proceedings of the XXVI Congreso de AESLA, Spain*, 3-5.
- Pecorari, D. (2003). Good and original: Plagiarism and patchwriting in academic second-language writing. *Journal of Second Language Writing*, 12, 317-345.
- Pennycook, A. (1996). Borrowing others' words: Text, ownership, memory and plagiarism. *TESOL Quarterly*, 30(2), 201–230.
- Petrić, B. (2007). Rhetorical functions of citations in high-and low-rated master's theses. *Journal of English for Academic Purposes*, 6, 238-253.
- Rayson, P. (2001). Wmatrix: A web-based corpus processing environment (Technical

- report). Lancaster: Lancaster University.
- Shaw, P. (1992). Reasons for the correlation of voice, tense, and sentence function in reporting verbs. *Applied Linguistics*, 13(3), 302-319.
- Shim, Eunsook. (2006). Use of citations in academic writing. *English Language Teaching*, 18, 157-178.
- Sun, Y. C., & Wang, L. Y. (2003). Concordancers in the EFL classroom: Cognitive approaches and collocation difficulty. *Computer Assisted Language Learning*, 16(1), 83-94.
- Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge, UK: Cambridge University Press.
- Thompson, P. (2000). Citation practices in PhD theses. In L. Burnard & T. McEnery (Eds.), *Rethinking language pedagogy from a corpus perspective* (pp. 91-101). Frankfurt: Peter Lang.
- Thompson, P. (2001). *A pedagogically-motivated corpus-based examination of PhD theses: Macrostructure, citation practices and uses of modal verbs*. Unpublished doctoral dissertation, University of Reading, Reading.
- Thompson, P. (2005). Points of focus and position: Intertextual reference in PhD theses. *Journal of English for Academic Purposes*, 4(4), 307-323.
- Thompson, P., & Tribble, C. (2001). Looking at citations: Using corpora in English for academic purposes. *Language Learning & Technology*, 5, 91-105.
- Thompson, G., & Ye, Y. Y. (1991). Evaluation in reporting verbs used in academic papers. *Applied Linguistics*, 12, 365-382.

APPENDIX

Journal Corpus

1. Bylund, E., Abrahamsson, N., & Hyltenstam, K. (2010). The role of language aptitude in first language attrition: The case of pre-pubescent attriters. *Applied Linguistics*, 31(3), 443-464
2. Kuppens, A. H. (2009) English in advertizing: Generic intertextuality in a globalizing media environment. *Applied Linguistics*, 31, 115-135.
3. Lee, J-H. (2010). A subject-object asymmetry in the comprehension of wh-questions by Korean learners of English. *Applied Linguistics*, 31(1), 136-155.
4. Lam, P. W. Y. (2010). Discourse particles in corpus data and textbooks: The case of well. *Applied Linguistics* 31(2), 260-281. A subject-object asymmetry in the comprehension of wh-questions by Korean learners of English. *Applied Linguistics*, 31(1), 136-155.

5. Shi, L. (2004). Textual borrowing in second language writing. *Written Communication, 21*, 171–200.
6. Lorenzo, F., Casal, S., & Moore, P. (2010). The effects of content and language integrated learning in European education: Key findings from the Andalusian sections evaluation project. *Applied Linguistics, 31*, 418-442.
7. Aylin, K., & Simon, M. (2011). Changes to English as an additional language writers' research articles: From spoken to written register. *English for Specific Purposes, 30*(20), 113-123.
8. Chou, M. (2011). The influence of learner strategies on oral presentations: A comparison between group and individual performance. *English for Specific Purposes, 30*(4), 272-285.
9. Chang, C. F. (2011). A corpus-based approach to online materials development for writing research articles. *English for Specific Purposes, 30*(3), 222-234.
10. Yamashita, J., & Jiang, N. (2010). L1 influence on the acquisition of L2 collocations: Japanese ESL users and EFL learners acquiring English collocations. *TESOL Quarterly, 44*(4), 647-668.
11. Hsu, W. (2011). The vocabulary thresholds of business textbooks and business research articles for EFL learners. *English for Specific Purposes, 30*(4), 247-257.
12. Thøgersen, J., & Airey, J. (2011). Lecturing undergraduate science in Danish and in English: A comparison of speaking rate and rhetorical style. *English for Specific Purposes, 30*(3), 209–221.
13. Kuteeva, M. (2010). Wikis and academic writing: Changing the writer–reader relationship. *English for Specific Purposes, 30*, 44-57.
14. Rubio, M. (2011). A pragmatic approach to the macro-structure and metadiscoursal features of research article introductions in the field of Agricultural Sciences. *English for Specific Purposes, 30*, 258-271.
15. Soler, V. (2011). Comparative and contrastive observations on scientific articles written in English and Spanish. *English for Specific Purposes, 30*, 124-137.

Examples in: English

Applicable Languages: English

Applicable Levels: Tertiary

Myung-Jeong Ha
Department of English Language & Literature
Sangmyung University
31 Sangmyungdae-gil, Dongnam-gu,
Cheonan, Chungnam, 330-720, Korea.
Tel: (041) 550 - 5153
E-mail: mjha@smu.ac.kr

Received 16 March 2012

Revised 7 May 2012

Accepted 19 May 2012

