A Corpus-based Contrastive Study on Reporting in English M.A. Theses

Jiang Yuechun，Hu Zhiqing

1 Introduction

Reporting, in a broad sense, means the account of what people say or do or think with the evaluation of the reporter or writer; narrowly, it refers to making references to previous research embedded in a continuous academic research article (Thomas and Hawes 1994:129). Whether the writers have used reporting appropriately or not has great impact on the quality of academic articles. However, it seems that reporting is also a source of considerable difficulty for most EFL writers (Cadman 1997, Thompson 2000).

Many researches have been done concerning reporting recently (Thompson 1994; Hyland 1999, 2002; Charles 1992; Thompson 2000; Okamura 2003; He 2000; Huang 2001). Roughly speaking, the previous studies of reporting take the following approaches, namely the stylistic approach (Leech and Short, 1981; Shen, 1991), the general linguistic approach, i. e. the lexical, syntactic and semantic approach (Swales, 1986, 1990), the pragmatic approach (Jia, 2000) and the functional approach (Thompson, 1994; Chen, 2006). However, to a large extent, the previous studies on reporting ignored the distribution of reporting in different discourses and the functional aspects of reporting. '

In this paper I investigate reporting in English M.A. theses written by Chinese Learners of English (CLE) and Native Speakers of English (NSE). I focus particularly on the frequency of the various structures and forms of reporting and how extensive differences can be revealed using a combination of both quantitative and qualitative data. The corpus contains 13 M.A. theses altogether, 6 of which, written by NSE, are randomly collected from Internet and 7 others written by CLE. It covers topics like language, communication and biology etc. The CLE's sub-corpus

contains 100,519 words, while that of NSE 100,218 words. After constructing the corpus, we read it line-by-line and inserted tags in each reporting phenomenon.

2 Reporting Structures

The reporting structures are various, consisting of finite verbs, present participles, past participles, reporting adjuncts, reporting nouns and reporting adjectives. Altogether, 598 reporting sentences in CLE's corpus and 767 sentences in NSE's one are extracted respectively, which indicates higher density of reporting in NSE's theses.

Table 2.1 displays the contrastive distribution of reporting structure. It's clear from the figure that finite verbs prevail over reporting sentences in both sub-corpora. Numbers of other reporting structures are relatively close. CLE's bar starts high but drops sharply, with NSE's bar remaining relatively stable, which suggests some reporting structures like reporting nouns, present participles, past participles occur in CLE's corpus in a very limited numbers. It's interesting to note that a similar number of reporting adjectives are adopted in their CLE and NSE's theses (Refer to table 2.1. for distribution of reporting structure). A closer investigation in the following will give us the detailed information.

TABLE 2.1 DISTRIBUTION OF REPORTING STRUCTURES IN CLE'S AND NSE'S

SUB-CORPORA

	CLE's corpus		NSE's corpus	
Item	Freq.	%	Freq.	%
Finite verbs	496	82.9	451	58.8
Reporting adjuncts*	59	9.8	158	20.6
Reporting nouns*	15	2.5	53	6.9

RVs in Present Participle*	11	1.8	59	7.7
RVs in Past Participle*	10	1.6	39	5.1
Reporting adjectives	7	1.2	7	0.9
Total*	598	100	767	100

Note: * significant difference at 0.05 level

2.1 Finite verbs

With both active and passive finite verbs included, finite verbs are investigated with great emphasis. It should be noted that all frequencies given in results are based on the counts of token occurrences. For example, for the occurrences of the finite verb '*suggest*', its variations like the past form or past participle '*suggested*' and the present participle '*suggesting*' are also counted, whether active or passive, only if these verbs are predicating words. What's more, the verbs under investigation are only finite verbs identified here, excluding verbs in past participle and present participle and those in reporting adjuncts.

The number of finite verbs ranks the first in both sub-corpora, for which this study lays great emphasis on them. The frequency, as high as 496 in CLE's corpus and 451 in NSE's corpus, with the percentage accounting for all reporting instances being 82.9% and 58.8% does suggest it's the most common reporting structure adopted by both CLE and NSE. This result is consistent with He's (2000) findings. In his research, the percentage of Finite verbs accounting for total reporting instances is 85.55%. However, the difference in percentage (82.9% in CLE corpus and 58.8% in NSE corpus) suggests that Finite verbs are CLE's favorite reporting structure, while NSE don't attach such great importance to this reporting structure as CLE do.

Examples are enormous, such as:

Esrock and Leichty (1998) **discovered** that of the selected Fortune 500 corporations, only 19% placed prominence on general community service or responsibility to the local community.

2.2 Reporting adjuncts

The reporting adjuncts are usually overlooked in traditional grammar, though it's easy to perceive their functions. In this thesis, reporting adjuncts include reporting adverbs, such as *'apparently', 'reputedly', 'allegedly', 'reportedly',* and *'supposedly'*, prepositional phrases, such as *'due to', 'according to'*，*'in accordance with',* infinitive phrase, like *'to quote',* subordinative finite clause, like *'as far as…is concerned', 'as…points out/said/admitted'*. It should be noted that if a reporting verb appears in reporting adjunct, as in the above example, it will no longer be considered as finite reporting verbs.

Reporting adjunct is the second favorite reporting structure by both CLE and NSE, with distinct frequency 59 to 158 in the two sub-corpora. Reporting adjuncts, as we mentioned before, contain various forms: reporting adverbs, prepositional phrases, infinitive phrase and subordinative finite clause.

Examples found in the corpus are various:

However, **as Martha Kolln points out** regarding Elbow's notion of the death-grip of grammar.

2.3 Reporting nouns

Reporting nouns, together with reporting adjuncts and reporting adjectives are completely new concepts proposed by Thompson (2000: 91-101). With a few exceptions, the majority of

reporting nouns are closely related to their relevant verbs; the that/whether/wh- clause after the nouns are the message carriers, giving complement information about the reporting content. More possible reporting nouns are: apology, admission, description.

15 in CLE's and 53 in NSE's corpus are found which enables us to confidently say there are significant differences between the two sub-corpora. CLE use much fewer reporting nouns than NSE do.

For instance:

There is general ***acceptance*** that good ICT programs can benefit participants in developing good relationships with hosts and completion of assignments rather than premature termination (Brislin and Yoshida 1994).

In this sentence, the reporting content is embedded in the word 'acceptance'. The majority of reporting nouns, with a few exceptions, are closely related to their relevant verbs; the that/whether/wh- clauses after the nouns are the message carriers, giving complement information about the reporting content. In this example, it's equal to say '***It's generally accepted*** that …'

2.4 Reporting verbs in present participle

Accordingly, present participles are counted as a form of reporting when they don't act as predicates in sentences.

Except for finite verbs, the RVs in present participle are far from being neglected. The frequency of RVs in present participle found in NSE's corpus is about five times as much as that in CLE's corpus, with a ratio of 59 to 11. For example:

Halliday and Hasan provide a comprehensive taxonomy of cohesion, ***classifying*** it into five types: reference, substitution, ellipsis, conjunction, and lexical cohesion.

2.5 Reporting verbs in past participle

Separate past participles are counted as a form of reporting when they don't act as predicates in sentences, most of which are complements to modify or to give further specific information.

The frequency of RVs in past participle in NSE's corpus is four times of the number of those in CLE's corpus, as 39 to 10. The numbers are small, since we have only 200,000 words in the corpus, but it doesn't mean the results aren't significant. For example:

One of the definitions most often referred to is the one ***provided*** by Tarone 1980, who considers communication strategies.

2.6 Reporting adjectives

This is also a new category suggested by Thompson. Those words are mostly used to report one's feeling. The common reporting adjective is '*so-called"*, in which reporting meaning is embedded.

There are a small number of reporting adjectives extracted in the corpus, such as *so-called,* and *aware*, with the amazingly same frequency of 7. For example:

Therefore, quite a number of teachers (Graham 1997) are ***aware*** of the need to boost students' confidence, both in terms of oral participation and confidence in their general linguistic abilities, and socialinguistic competence.

There are 7 reporting adjectives in both sub-corpora. The small number of reporting adjectives may be due to the function of reporting adjective itself. Its functions are, according to Thompson (2000: 86), to demonstrate the purpose of words, or reflect content of real conversation; while during research article writing, authors pay more attention to the information conveyed, not

the way to convey the information. That may cause the small number of reporting adjectives. The identical 7 reporting adjectives display that both CLE and NSE do not regard them as a common method to report.

The analysis of reporting structure is revealing. Interesting variations in the reporting structures have been found between the CLE and NSE sub-corpora. NSE appear to employ a much larger number of reporting sentences in their theses, with more dispersive reporting number in each reporting structure. The relatively high use of various reporting structures may result from their better mastery of the language skills and awareness to use diverse reporting structures to avoid monotony.

3 Reporting Forms

The reporting form is a functional category, referring to obscuring or revealing the writers' presence in their texts. There are four categories identified in the study: integral and non-integral reporting, non-citation and stative reporting.

3.1 Integral and non-integral reporting

If the reporting is in the form of a name followed by year number, typically the name will be incorporated into the sentence as an integral part of the syntax of the sentence, and will not be separated by brackets (Thompson 2000), it is called integral reporting. Obviously, integral reporting appears within sentences.

For example:

Kent and Taylor (1998) *suggest* that dialogic communication involves a relational interaction.

On the contrary, non-integral reporting is separated from the sentence by brackets and it plays no explicit grammatical role in the sentence. Reporting can also take the form of a number (rather than name and year) (Thompson 2000), telling the reader where the information (verbal or numerical) or idea comes from. The function of the reporting is that of attribution.

For example:

Additionally, research has found that 74% of consumers said that they were prepared to switch brands if a similar brand was associated with a worthy cause (Adkins & Kowalska 1997; New Zealand Marketing Magazine 2000, as cited in McAlister & Ferrell 2002).

TABLE 3.1 FREQUENCY OF REPORTING FORMS IN EACH SUB-CORPUS

Forms	CLE's corpus		NSE's corpus	
Integral	268	56.8	161	36.9
Non-integral	99	21.0	143	32.8
Non-citation	80	16.9	98	22.5
Stative	25	5.3	34	7.8
Total	472	100	436	100

Note: * significant difference at 0.05 level

It's clear from the table that there is much more integral reporting in CLE's corpus, with 268 in CLE's corpus against 161 in NSE's corpus. On the contrary, as to the non-integral reporting, the number is 204 in CLE's corpus against 275 in NSE's corpus.

The choice between integral or non-integral reporting indicates different focuses of writers. For those to use more non-integral reporting, writers tend to focus on previous findings or

suggestions rather than on the researchers that have made the findings or suggestions. On the other hand, more integral reporting indicates there is more attention paid to the individuals who have developed approaches, expressed equations, or who have articulated complex models, so they appear as actors within sentences (Thompson 2000). Therefore, it's safe to say that in the NSE's theses, writers put more focuses on the findings, or suggestions that have been carried out; on the other hand, CLE pay more attention to the persons who have put forward the findings or suggestions.

3.2 Non-citation and stative reporting

In the reporting form of non-citation, occurrences of a name in the text did not appear as a citation (i.e. no year, or page, reference attached to the name). The exception to this was instances where the name was used to identify a particular theorem, model, law, or other such commonly recognized construct. These 'non-citations', occur, of course, after the researcher has already been reported. (Thompson 2000)

For example:

Daugherty (2000) suggests that contributions of employee time and talents provide more public relations and marketing benefits. She *notes* that forms of voluntarism are seen as more sincere and provide additional benefits for employees who share a unified goal outside of the workplace.

The fourth reporting form is stative reporting, which is general. Also no year, or page, reference attached to the name, but it refers to previous research in a general and independent way. By 'independent', we mean there are no related reporting previously mentioned.

For example:

Finally, some researchers simply described a group of writers' products to aid in pedagogy.

The number of 80 to 98 in CLE's and NSE's sub-corpora tells that there is not much difference of the non-citation phenomena. In this reporting, there is no specific source for the statement, without year, or page, reference attached to the name. It refers to previous research in a general way, or a convention or an established fact or a theory. But it easily leads to some overgeneralization limited by one's knowledge or range of research.

The number of stative reporting in CLE's corpus is a little smaller than the one in NSE's corpus, with 25 to 34, but no significant difference is identified. Both numbers are small, which might suggest CLE and NSE have similar understanding of this type of reporting. The similar number signifies that CLE postgraduates are aware of or subconscious of the function of stative reporting, since stative reporting is general and independent and it doesn't provide the information source as most reporting cases do.

4 Qualitative Analysis on Functions of Reporting

The previous study has indicated the differences and similarities of reporting in the corpus. However, what we've done is restrained to the number of words used; the pragmatic usage, i.e. how well they are used, is not mentioned. For an adequate analysis and description of reporting, especially on how RVs function in their theses, the use of qualitative is also crucial, which can only be revealed by additional detailed examination of the context.

During the line-by-line reading of the corpus, we discover that there are some cases that demonstrate the uncertainty in the using of reporting for some CLE postgraduates. For consideration of length, I'd pick some paragraphs from CLE's corpus and look closely at those

reporting sentences employed.

Here are some continuous paragraphs taken from CLE's corpus in which so many 'argue's there catch our eyes. The original paragraphs go like this:

In the past thirty years a substantial amount of research has accumulated regarding the nature and prevalence of communication apprehension (CA). Defined by McCroskey (1977a) as "the fear or anxiety associated with either real or anticipated interaction with others", several researchers **argue** [1] that no other variable in communication research has received as much attentions (e.g., see Levine & McCroskey 1990; Lustig &Andersen 1991; Payne & Richmond 1984). Other constructs related to communication apprehension have been studied extensively as well…

Normative data indicates that approximately 15 to 20 percent of the United States population experiences high levels of trait CA: that is, anxiety with either real or anticipated interaction with others. McCroskey & Richmond (1996) **argue** [2] that virtually 100 percent of the population experiences one of the four contextual types of CA at some point. Buss (1980) **argues** [3] that some of the salient situational features leading to increased anxiety include novelty, unfamiliarity, and dissimilarity … Based on Buss's (1980) criteria, initial interaction with someone, or interacting with strangers, may produce heightened anxiety in persons. Berger and Calabrese (1975) **argue** [4] that "whenever two people come together and interact for the first time, they have a very limited amount of information about each other. In such circumstances, considerable uncertainty exists. High levels of uncertainty lead to increased anxiety". Berger and Calabrese (1975) **argue** [5] that "in such situations the primary goal of the interactants is to reduce uncertainty and to reduce uncertainty and to

increase the predictability about the other…

One type of communication situation that is potentially replete with novelty, unfamiliarity, dissimilarity, and uncertainty is intercultural communication. Gudykunst and Kim (1997) **argue** [6] that when individuals are confronted with cultural differences they tend to view people from other cultures as strangers. Strangers are unknown people who are members of different groups. Anyone entering a relatively unknown or unfamiliar environment falls under the rubric of stranger. In their conceptualization, Gugykunst and Kim (1997) content [7] that interaction with people from cultures other than our own tend to involve the highest degree of strangeness and the lowest degree of familiarity. Thus, there is greater uncertainty in initial interaction with strangers than with people who are familiar. In such circumstances not only is uncertainty high but so is anxiety. According to Gudykunst and Kim (1997) [8], actual or anticipated interaction with members of different groups (e.g., cultures different from our own to anxiety).

This piece of writing represents two typical weak points of CLE in using reporting: inappropriate function of some RVs and overuse of a few RVs.

The frequent word here is 'argue'. By using the word 'argue', from dictionary explanation, the speaker/author is trying to persuade others by reasoning and giving evidence when there are other people holding opposite points of view. For example, '*I argued with her for a long time, but she refused to listen to reason.*'

Hence, we can see inappropriate usage of 'argue' here besides obvious grammatical mistakes. Example [1] is a little confusing when '*several researchers argue that no other variable in communication research has received as much attentions (e.g., see…*' The literal meaning of this

sentence indicates that some said there were some variables had received more attention than CA, while these researchers didn't agree with them; but judging from the context, we can say that the author actually means that 'the nature and prevalence of CA' has received much attention, as is displayed in those researchers' work.

In Example [3], those salient situational features brought forward by Buss never occurred previously, neither its opposite findings did. So it is with Example [4], [5] and [6]. The writer didn't make reporting verbs serve his/her purpose well, employing one verb while actually meaning another inferred from the context.

The only exception is Example [2], in which the two researchers express different opinions from previous ones and the word 'argue' is just the right word to describe this situation. The lacking accuracy of the word 'argue' reflects the weak points of some CLE in the mastery of verbs.

From the context, in Example [1], [3], [4], [5] and [6], reporting verbs such as 'suggest', 'present' or 'state' could be used instead of 'argue'.

Another distinguished feature of these paragraphs is the excessive use of the word 'argue'. This individual writing style exerts great influence on the results when we are counting the tokens and frequencies of RVs. There are seven RVs altogether, among which 6 are 'argue'. The excessive use of some verbs reflects lacking style awareness to avoid monotony in writing. There are some other cases with too frequent 'find' or 'analyze'.

This inadequate variety is also represented in reporting structures. In this piece of continuous writing, there is only one reporting adjunct (example [8]), with the other 7 occurrences all RVs. The relatively insufficient employment of various reporting structures is also reflected in the total number in CLE's corpus.

It's true that each writer has some favorite reporting structures or reporting verbs, but we find

particularly in CLE's corpus, the individual style varies from one postgraduate to another. A certain passive or active sentences, for example, are used with extremely high frequency in this corpus, thus adding unreliability to the results. A redeeming way for future research is suggested to investigate each article instead of the whole corpus so that we can have a clearer view of the distribution of each element under discussion, and reduce those factors caused by writers' particular preference.

These differences could result from students' different proficiency levels of the language itself, awareness of writing skills, transference of Chinese in the second language learning, different thinking patterns and cultural factors (Kaplan 1966), or different epistemologies in which these students have been trained (Cadman 1997). It is confirmed that EFL learners' native language and culture have great impact on their thesis/dissertation writing (Yu 1998; Shaw 1992). Moreover, the survey results suggest that this language transfer may even exist at the advanced level also, after the writer has already acquired a high level of English language proficiency. Therefore, long-term and in-depth investigations are needed to further determine how much and how language and cultural differences impact CLE's thesis writing.

Generally speaking, the difference between the CLE's and NSE's theses is significant and it is suggested that CLE can improve their writing in the following ways: flexibility and variety of reporting methods, awareness of writing skills, and a more objective reporting.

5 Conclusion

The research finds that there are considerable similarities of usage and remarkable differences as well in reporting between the CLE's and NSE's sub-corpora. Specifically, in terms of reporting structure, NSE's theses possess higher density of reporting; finite verbs prevail over reporting sentences in both sub-corpora, but some reporting structures like reporting nouns, present

participles, past participles occur in CLE's corpus in a very limited number.

Talking about the reporting forms, CLE prefer non-integral reporting, while NSE prefer integral reporting, which indicates CLE pay more attention to previous findings or researches while NSE concentrate on the one who makes the research, which suggest a more 'impersonal' construction in NSE's corpus. Case study also shows better flexibility and skillfulness are reflected in the use of reporting in NSE's theses and CLE are not skillful enough to express their ideas when arguing for their own benefit.

The pedagogical implications are that this paper helps teachers to know where to put their focuses in the teaching of academic writing. It also helps writers to improve their awareness of reporting in appreciating and writing academic articles.

References

[1]   Cadman, K. (1997) Thesis writing for international students: A question of identity? *English for Specific Purposes* 16, No. 1, 3-14.

[2]   Charles, M. (1992) Evaluation in report sources: uncovering disciplinary differences in theses. *Now you see me, now you don't: the author's voice in academic research articles*. Oxford University Language Centre.

[3]   Chen, M. F. (2006) *Reporting in Literature Reviews of English Dissertations* Xiamen: Xiamen University

[4]   He C. W. (2000) *A Corpus-based Study of Reporting in Academic Research Articles*. Wuhan: Department of Foreign Languages, Huazhong University of Science and Technology.

(贺灿文，2000，《英语学术论文中的语料库转述语研究》。武汉：华中理工大学。)

[5] Huang, Y. J. (2001) Subject-predicate Collocation in reporting sentences. *Foreign Language Education.*

（黄谊军，2001，综述句中的主谓搭配。《外语教育》。）

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

[7] Hyland, K. (2002) Authority and invisibility: Authorial identity in academic writing.Journal of Pragmatics, 34: 1091-1112.

[8] Jia, Z. H. (2000) Reported speech and its pragmatic function. *Foreign Language,* 2.

（贾中恒，2000，转述语与其语用功能。《外国语》，2。）

[9] Kaplan, R. (1966) Cultural thought patterns in inter-cultural education. *Language Learning* 16, 1–20.

[10] Leech, G. N. and M. Short. (1981) *Style in Fiction*. London: Longman.

[11] Okamura, A. (2003) How do British and Japanese scientists use 'we' and verbs in biology, chemistry and physics papers. *The Economic Journal of Takasaki City University of Economics.*

[12] Shen, D. (1991) Different speech expressions in novels. *Foreign Language Teaching and Research*, 1.

（申丹，1991，小说中的不同转述语，《外语教学与研究》，第1期。）

[13] Shaw, P. (1992) Reasons for the correlation of voice, tense, and sentence function in reporting verbs. *Applied linguistics*, 13, issue, 3.

[14] Swales, J. (1986) Citation analysis and discourse analysis. *Applied Linguistics*, 7/1, 39-56.

[15] Swales, J. (1990) *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press. 150-152

[16] Thomas, S. & Hawes, T. P. (1994) Reporting verbs in medical journal articles. *English for Specific Purposes*, 13/2, 129-148.

[17] Thompson, G.. (1994) *Collins Cobuild English guides 5: Reporting*. London: Harper Collins Publishers.

[18] Thompson, P. (2000) Citation practices in Ph. D. theses. In L. Burnard & T. McEnery (Eds.), *Rethinking language pedagogy from a corpus perspective*. Frankfurt: Peter Lang. 86-101.

[19] Yu R. D. (1998) Non-native Graduate Students' Thesis/Dissertation Writing in Science: Self-reports by Students and Their Advisors from Two U.S. Institutions. *English for Specific Purposes* 17, No. 4, 369–390.