Use of *That*-Clauses After Reporting Verbs in Asian Learners' Speech and Writing: Frequency, Verb Type, and *That*-Omission

Shin'ichiro Ishikawa[1]

[1] Kobe University, Japan
iskwshin@gmail.com

Abstract

That-clauses after reporting verbs (VTHAT) are widely used in L2 and L1 English. Previous studies have examined their frequency, common reporting verbs, and omission of the complementizer, but how varied learners use VTHAT in their writing and speech and how they differ from native speakers in usage of VTHAT has not been wholly elucidated. Therefore, using the International Corpus Network of Asian Learners of English (ICNALE), we compared the uses of VTHAT by six groups of Asian learners of English (ALE) at different L2 proficiency levels and English native speakers (ENS). Our analyses have revealed that ALE use VTHAT less often than ENS, omit the complementizer more often both in speech and writing, and tend to use reporting verbs such as "think," "believe," "agree," and "know."

1 Introduction

That-clauses after reporting verbs (VTHAT) are present in varied genres of L1 and L2 English texts (e.g., I think that Mike should do it). The subject of the main clause often refers to the human participant, the reporting verb, namely, the lexical controlling verb presents the type of reporting (speech or thought), and the *that*- complement clause, placed in post-predicate position, presents the reported speech, thought, attitude, or emotions of humans (Biber, Johansson, Leech, Conrad, and Finegan, 1999, p. 660).

Although the literature has touched upon several facets of the usage of VTHAT, how they are used in speech and writing by learners as well as native speakers has not been wholly described. Thus, as Biber and Reppen (1998) emphasize, many of the existing reference books and English as a Second Language/ English as a Foreign Language (ESL/EFL) grammars do not necessarily offer appropriate answers to basic questions such as "Which structural types of complement clauses are common and which are rare?" "Are these structures found primarily in speech or writing?" and "Are any particular verbs especially common controlling complement clauses?" In the current study, therefore, we aim to

see how varied Asian learners of English (ALE) and English native speakers (ENS) adopt VTHAT in their spoken and written outputs.

2 Literature Review

Previous studies have analyzed VTHAT in L1 and L2 English mainly in terms of (1) overall frequency, (2) omission of *that* as a complementizer, and (3) types of major reporting verbs.

2.1 VTHAT Seen in L1 English

Concerning (1), Biber and Reppen (1998) analyzed a large-scale corpus and showed that the per-million-words (PMW) frequencies of VTHAT were 6,000+ in conversations, 5,000+ in fiction, 4,000+ in news, and 1,000+ in academic prose. They also added that in conversations, fiction, and news, VTHAT occur more often than other verb constructions including *to/ ing/ wh*-clauses (p. 150).

Biber (2006) analyzed a corpus of spoken and written texts concerning academic settings and showed that the PMW frequencies of VTHAT were 8,000+ in conversations during office hours or in study groups, 7,000+ in conversations in labs or at service encounters as well as in classroom teachings, 3,000+ in textbooks and course packs, and 2,000+ in syllabi and institutional writing (p. 81). This analysis implies that VTHAT are essentially speech- or conversation-oriented.

Then, concerning (2), Biber et al. (1999) suggested that the rates of *that*-omission (e.g., I think ~~that~~ Mike should do it) are 85%, 59%, 27%, and 6% in conversations, fiction, news, and academic prose respectively (the ratios shown above have been estimated by the author from Figure 9.7 on p. 680). The *that*-omission rate clearly increases as the genre becomes more colloquial and speech-oriented. Thus, *that*-omission forms are more common in conversations and fiction, while *that*-retention forms are more common in news and academic prose. Biber et al. (1999) add that the complementizer *that* tends to be omitted when it is preceded by "think" and "say," co-referential subjects appear in main or *that*-clauses, and personal pronoun subjects occur in *that*-clauses; it tends to be retained when coordination occurs in *that*-clauses, the main clauses include passive voice, and noun phrases intervene between main clauses and *that*-clauses (pp. 680–682). According to Biber (2006), office-hour conversations and classroom teaching at colleges are characterized particularly by "I think/ mean/ guess ~~that~~" and "we/ you know that" forms respectively.

Finally, concerning (3), Carter and McCarthy (2006) analyzed large corpora and listed 68 of the most common reporting verbs with *that*-clauses as the direct object.

accept, admit, agree, announce, argue, assume, believe, bet, check, claim, comment, complain, conclude, confess, confirm, consider, decide, deny, discover, doubt, expect, explain, feel, find, forget, gather, guarantee, guess, hear, hint, hold, hope, imagine, imply, infer, insist, know, learn, mean, mention, notice, predict, presume, pretend, promise, protest, prove, realize*, recall, reckon, recognize*, remark, remember, repeat, reply, report, say, see, show, state, suggest, suppose, suspect, swear, think, understand, warn, write   (*-ise in the original)

Table 1: Common reporting verbs (Carter and McCarthy, 2006, p. 511)

Biber et al. (1999) also conducted an extensive corpus analysis and listed 153 major reporting verbs, classified into mental verbs, speech act verbs (e.g., say, tell), and other (non-verbal) communication verbs (e.g., show, prove, suggest). Mental verbs comprise frequently occurring cognition verbs (e.g., think, know) as well as less frequent emotive/affective verbs (e.g., hope, wish) and verbs concerning receptive processing of communication (e.g., read, hear; p. 661, 666). The table below lists 39 "notably and relatively common" reporting verbs belonging to three semantic types. The frequencies are given in per-million words (PMW).

Types	Notably common (100+)	Relatively common (20+)
Mental	think (600+), know (400+), see/ find (200+), believe, feel, guess (in American English)	assume, conclude, decide, doubt, expect, hear, hope, imagine, mean, notice, read, realize, recognize, remember, suppose, understand, wish
Speech act	say (600+)	admit, agree, announce, argue, bet, insist
Communication	show, suggest	ensure, indicate, prove

Table 2: Common reporting verbs belonging to three semantic types (Biber et al. 1999, pp. 663–666)

As shown in the table above, among varied VTHAT forms, "X think that," which conveys a sense of uncertainty, "X know that," which expresses a sense of certainty, and "X say that," which quotes someone's utterance are said to be especially frequent.

In addition, Biber et al. (1999) suggested that the frequencies of the most common reporting verbs may change across different text genres.

Genres	Conversation	Fiction	News	Academic
Verbs	think (2000+) say (1200+) know (600+) guess (400+)	think (1000) say/ know (800) see/ find/ feel (200+)	say (1600+) believe/ think (200+)	show (200+) suggest/ say (200)

Table 3: Common reporting verbs in different genres (Biber et al. 1999, pp. 668–669)

In conversations as well as fiction, "X think that" is the most common form, which often conveys a speaker's uncertainty about the reported event. Meanwhile, in news and academic prose, "X say that" and "X show that" are used most often, both of which present the fact in an objective way.

2.2 VTHAT Seen in L2 English

Concerning (1), Biber and Reppen (1998) analyzed 280 thousands-tokens of essays written by French, Spanish, Japanese, and Chinese learners and reported that the PMW frequencies of VTHAT are 7,000+ for Spanish learners, 6,000+ for Japanese learners, and 5,000+ for Chinese and French learners. This means that the average frequency of VTHAT in L2 English is approximately 5,800+, which is higher than that in four genres of L1 English (4,000+; Biber, et al., 1999).

Granger (1998) analyzed an international learner corpus and reported that a certain set of VTHAT forms, which she calls sentence-builders for stating the discourse purposes, are extremely overused by European learners of English. For example, learners use "[generalized pronouns] + [modals] + say that" forms 18 times (4 for native speakers vs. 75 for learners) as much as native speakers, and "I think that" forms 24 times (3 vs. 72). They also frequently use "[generalized pronouns] + [modal] + notice that" and "[generalized pronouns] + [modal] + not forget that" forms, both of which are not observed in native speakers' essays (pp. 154–155).

Learners' overuse of "I think (that)" forms has attracted a lot of attention from second language acquisition (SLA) scholars. For instance, Sakaue (2013) reported that "I think that" is the third most frequently used trigram in English essays written by Japanese college students during the semester. Tono (2004) analyzed transcripts of Japanese learners' utterances in oral proficiency interviews and mentioned that "I think" is included in the top eleven bigrams used by intermediate and upper-level learners (Level 4–9; p. 110). József (2001) analyzed Hungarian college students' essays and reported that learners use "I think" seven times as much as native speakers (p. 128). Chen (2015) analyzed

Taiwanese learners' essays and concluded that they have a tendency to overuse "I think," which the author says is essentially a spoken discourse marker.

Then, concerning (2), Biber and Reppen (1998) illustrated that the *that*-omission rates are 45% for Spanish learners, 51% for Chinese learners, and 55% for French and Japanese learners (the ratios have been estimated by the author from Figure 11.3 on p. 154). The average is 51%, lower than the ratio in L1 conversations (85%) and fiction (59%), but higher than that in L1 news (27%) and academic prose (6%).

Finally, concerning (3), Biber and Reppen (1998) listed the reporting verbs most frequently used by different learner groups.

L1	French	Spanish	Chinese	Japanese
Verbs	think (1500)	think (2100)	think (1300)	think (2100)
	hope (1200)	know (1100)	hope (800)	hope (800)
	know (700)	say (900)	know (600)	say (600)
	say (500)	hope (800)	say (300)	know (400)

Table 4: Common main verbs used by different learner groups (Biber and Reppen 1998)

Note that "think," "know," and "say" characterizing L1 conversations and fiction are used frequently in a common way, while "show," frequently used by native speakers in writing, is not observed here. This may show that learners' essays are somewhat akin to L1 speech. Another interesting fact is that many learners frequently use "hope," which Biber and Reppen regard as one of the features characterizing L2 English.

3 Research Design

3.1 Aim and Research Questions

Although previous studies have revealed several noteworthy facts about the use of VTHAT in L1 and L2 English focusing mainly on (1) overall frequency, (2) *that*-omission, and (3) types of major reporting verbs, they are based on the analyses of several independent corpora that have not necessarily been designed for mutual comparison. In addition, they hardly consider learners' L2 proficiency levels. These points might influence reliability and replicability of the reported findings.

Therefore, in the current study, we use a newly developed international learner corpus to reconsider three basic facets of VTHAT use by ENS and six groups of ALE in China (CHN), Indonesia (IDN), Japan (JPN), Korea (KOR), Thailand (THA), and Taiwan (TWN) at different L2 proficiency levels, and thus, raise three research questions.

RQ1 How often do learners and native speakers use VTHAT in their speech and writing?
RQ2 How often do learners and native speakers omit *that* as a complementizer in their speech and writing?
RQ3 What type of reporting verbs do learners and native speakers use in their speech and writing?

3.2 Data

The current study uses the International Corpus Network of Asian Learners of English (ICNALE) (Ishikawa, 2013; Ishikawa, 2014). The ICNALE collects 10,000 samples of monologue speech and essays produced by 3,550 EFL/ ESL college students in 10 countries and areas in Asia as well as 350 ENS including college students, ESL teachers, and others. All the learners are classified into four

proficiency levels (A2, B1_1, B1_2, and B2+) linked to the Common European Framework of Reference, based on their scores in English proficiency tests such as TOEFL and TOEIC or English vocabulary size test.

All the participants, including learners and native speakers, are given two common topics— "It is important for college students to have a part-time job" (PTJ) and "Smoking should be completely banned at all the restaurants in the country" (SMK)— and required to write 200–300 word essays or to make 60-second speeches. In both cases, they are told to show clearly whether they agree or disagree with the given topics. While writing the essay, participants read the topics and write in MS Word without using dictionaries or other reference books. While producing their speech, participants make a phone call to the ICNALE speech data collection system, hear the topics, and then make impromptu (no scripts) speeches over the phone after a short preparation period. Their speeches are recoded and transcribed by professional transcribers.

Although the ICNALE includes data of six EFL learner groups and four ESL learner groups, we limit ourselves to discussing EFL learners only. The table below shows the size of each submodule used for the current analysis.

S/W	Level	CHN	IDN	JPN	KOR	THA	TWN	ALE	ENS	All
S	A2	6119	10727	7135	1671	688	5637	31977		
S	B11	20893	13948	12335	4202	6565	15036	72979		
S	B12	34826	14734	12014	11256	8872	9080	90782		
S	B2	4427	1282	9845	12410	1723	6759	36446		
S	All	66265	40691	41329	29539	17848	36512	232184	93459	325643
W	A2	33961	14798	68701	33286	53683	12905	217334		
W	B11	110631	37549	79820	26993	80990	40087	376070		
W	B12	52193	39622	22435	40261	45986	28549	229046		
W	B2	6590	1544	8550	36343	1016	11055	65098		
W	All	203375	93513	179506	136883	181675	92596	887548	90352	977900
SW	All	269640	134204	220835	166422	199523	129108	1119732	183811	1303543

Table 5: The number of words in the ICNALE submodules used for the analysis

3.2.1 Target Forms

Although VTHAT has varied formal variants, we examine only the case matching the pattern such as [noun OR personal pronoun] + [verb] + that + [noun OR personal pronoun]. Inappropriate samples (e.g., "It is/ seems that S+V") are manually excluded. The forms without *that* and those with intervening elements (e.g., "I *always* think that S+V," "I *am* thinking that S+V") are not considered here. As the target analyzed here is clearly narrower in a range than that discussed in previous studies, direct comparison of frequencies should be avoided.

3.3 Methodology

For RQ1 (overall frequency), the PMW frequencies of VTHAT are investigated and mutually compared. When conducting contrastive analyses, we discuss not only six Asian learner groups but also ALE in general. ALE frequency is the average of frequencies in the subcorpora of six learner groups.

When investigating RQ2 (*that*-omission), we compare the number of occurrences of VTHAT with and without *that* as a complementizer, focusing on the most common forms: "I + think + [that OR zero] + [noun OR personal pronouns]."

Concerning RQ3 (types of reporting verbs), we examine (i) the variety of reporting verbs, (ii) types of major reporting verbs, and (iii) the relation between learners/ native speakers and use of major reporting verbs. We discuss (i) in terms of the number of verb types and Herdan's C value (natural log of number of unlemmatized types divided by natural log of number of tokens), which is said to be a

more stable index for lexical variety than a traditionally used type/token ratio (TTR). In both cases, we deal with unlemmatized verb types. For (ii), the top five reporting verbs, their semantic types (mental, speech act, or communication), and the ratio of the three semantic types are compared. For (iii), we conduct a correspondence analysis to see the relationships between spoken and written texts by learners and native speakers (Item 1) and reporting verbs (Item 2). Item 1 includes 14 kinds of texts (speech and writing by six learner groups as well as native speakers), and Item 2 includes 17 high frequency reporting verbs occurring in more than six of 14 texts (agree, know, say, think, believe, disagree, said, feel, knows, means, shows, thought, claim, hope, says, find, realize). Affinities between items are shown on the scatter plot where dimension 1 (Z1) and dimension 2 (Z2) are regarded as a horizontal and vertical axis respectively.

4   Results and Discussions

4.1   RQ1 Overall Frequency

VTHAT frequencies in speech and writing by learners and native speakers, and how they change according to increase in learners' L2 proficiency levels (A2, B1_1, B1_2, and B2+) are shown in the figures below.
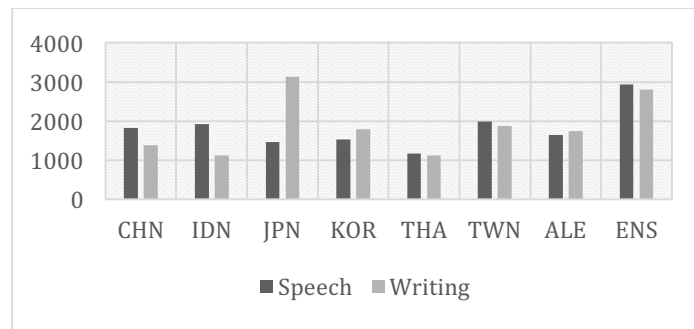

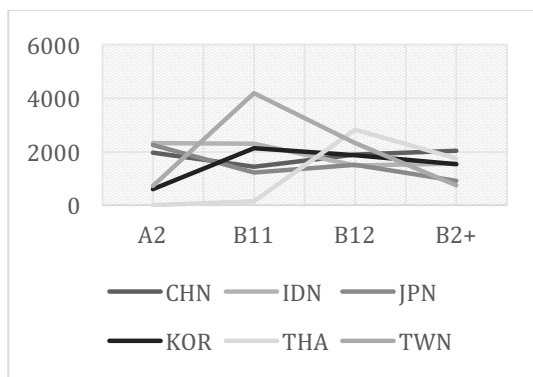**Fig. 1** PMW frequencies of learners and native speakers

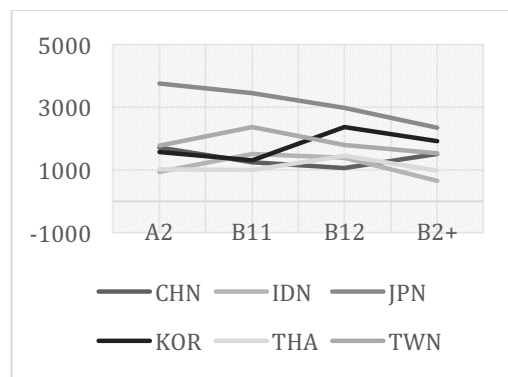**Fig.2** PMW frequencies in learner speech

Fig. 3 PMW frequencies in learner writing

Previous studies have suggested that VTHAT frequency in L1 conversations (6,000+) is higher than in writing (1,000–5,000+), and frequency in learners' writing (5,000–7,000+) is, in many cases, higher than in native speakers' writing and conversations (1,000–6,000+). In addition, learners are said to use particular types of VTHAT 18–24 times as often as native speakers do. However, our analysis based on the controlled corpus reveals new facts about the frequency of VTHAT used by learners as well as native speakers.

First, as shown in Fig. 1, native speakers use VTHAT almost equally often in speech (2,932) and writing (2,800). It is true that the frequency is somewhat higher in speech than in writing as the literature suggests, but the difference is rather small. On comparing the spoken and written monologues, the frequency of VTHAT seems to be almost the same. Higher frequency of VTHAT in conversations reported in previous studies might be attributed not to the difference between speech and writing but to the difference between dialogues and monologues.

Second, learners generally use VTHAT much less often than native speakers do in speech (1,635) and writing (1,735), which does not support the findings of previous studies (Granger, 1998). However, the frequency significantly varies across different learner groups. For example, Thai learners use a relatively small number of VTHAT in speech and writing, while Japanese learners use VTHAT extraordinarily frequently in writing. Japanese learners' overuse of VTHAT seems to be caused by too frequent use of the "I think that" and "I agree that" forms.

> *I agree that* it is important for college students to have a part-time job... Therefore *I agree that* college students have a part-time job, but *I think that* students who don't have a ability that they can take care of themselves should not have it.   (W_JPN_PTJ_092_A2_0)

> …*I think that* doing part-time job is the best way for them to change free time into useful time and to make money…. *I think that* we should do many kinds of part-time job for 4 years in college…. So, *I think that* we need do many part-time job to cope with anything easily….
>   (W_JPN_PTJ_084_A2_0)

Note that most of the expressions "I think/agree that" seen in the writing above occur before a writer makes some kind of strong claim ("it is important," "should not...," "… is the best way," "we should do…," and "we need [to] do…"). Japanese writers seem to try to distance themselves from what they write about and soften their claims by adopting VTHAT. Interestingly, Japanese learners do not overuse VTHAT too much in speech, where they presumably do not have enough time to control stances in such a delicate way.

Third, as shown in Fig. 2, the VTHAT frequencies do not necessarily correlate with learners' L2 proficiency levels. However, the frequency in writing by Japanese learners shows a consistently decreasing trend, meaning that the deviant overuse of VTHAT mentioned above is characteristic not of Japanese learners in general but of Japanese learners at novice or lower intermediate levels. In addition, all the learners except for the Taiwanese use fewer VTHAT as their proficiency level increases from B1_2 to B2+. Higher proficiency may be characterized by less dependence on VTHAT. Considering that learners generally use VTHAT less often in comparison to native speakers, their development in proficiency seems to mean estrangement from native speakers, not access to them.

4.2 RQ2 *That*-omission

That-omission rates in speech and writing by learners and native speakers, and how they change according to the increase in learners' L2 proficiency levels (A2, B1_1, B1_2, and B2+) are shown in the figures below.
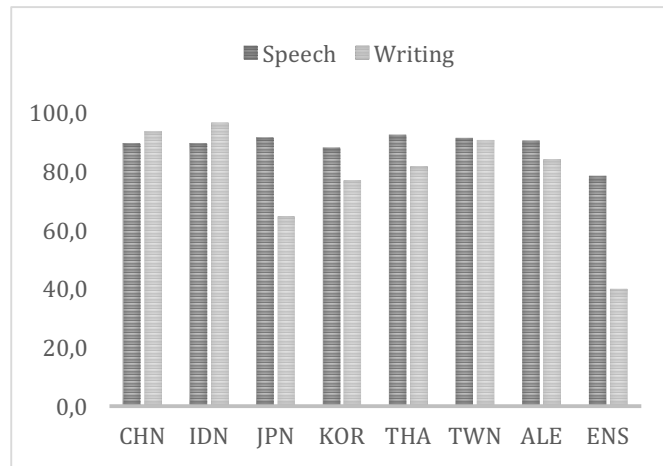
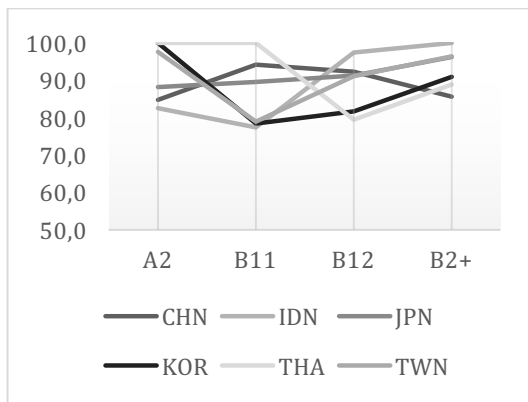**Fig. 4**: *That*-omission rates of learners and native speakers (%)

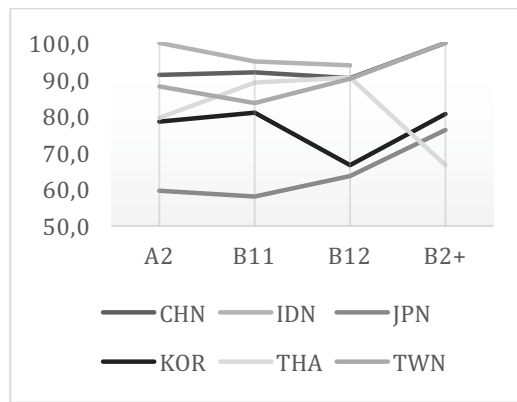**Fig. 5**: *That*-omission rates in learner speech   **Fig. 6**: *That*-omission rates in learner writing (NB: No hits for Indonesian learners at B2+)

Previous studies have reported that *that*-omission rates are 85% in L1 conversations, 6–59% in different genres of written texts, and 45–55% in L2 writing. Our analysis focusing on a typical structure ("I think (that) S+V") reveals several new facts.

First, as shown in Fig. 6, the *that*-omission rates are 78.1% in speech and 39.7% in writing for native speakers, roughly matching the previous findings. Note that the omission rate is higher not only in the dialogue-based conversations discussed in the literature but also in the spoken monologue analyzed in the current study.

Second, the average omission rates are 90.1% in speech and 83.8% in writing for learners. These rates are clearly higher than those for native speakers, meaning that learners omit complementizers more often than native speakers do. The difference in the omission rates between speech and writing is much smaller for learners, which may imply that learners are less sensitive to the stylistic or formal difference between speech and writing.

> ... *I think* you shouldn't work part time because your parent can give all of your daily need... But if your parents have a little bit finance problems *I think* you must have a part time job... *I think* it is depend in the college student situation... But if you study in Indonesia *I think* you shouldn't have a part time job because the salary is very cheap and can't meet your daily need.

(W_IDN_PTJ_147_B1_2)

Most of native speakers and some learners consistently retain or omit complementizers as shown above, while other learners seem to choose retention or omission rather on an ad-hoc basis.

I think that it is important for college students to have a part-time job... But *I think that* colleges are not the place which students only study... So *I think* it is important for us to have a part-time job...

(W_JPN_PTJ_185_A2_0)

Another important finding to be noted here is that the omission rate is the lowest in writing by Japanese learners, which may partly explain why Japanese learners use VTHAT extraordinarily often in writing compared to other Asian learners.

Third, as illustrated in Figs. 5 and 6, the *that*-omission rates change randomly for novice and lower intermediate learners, while they begin to increase in speech and writing for upper intermediate and advanced learners, except for Japanese and Thai learners. Most Asian learners omit the complementizers more often as their proficiency level goes up. Considering that learners generally omit *that* more often in comparison to native speakers, their development in proficiency seems to mean estrangement from native speakers here as well.

4.3   RQ3 Types of Major Reporting Verbs

4.3.1 Variety in the types of reporting verbs

The number of verb types and Herdan's C values, both of which represent the degree of variety in reporting verbs, are shown below.
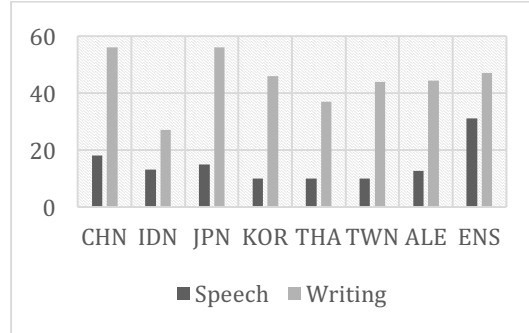

Fig. 7: Number of unlemmatized types of reporting verbs used by learners and native speakers
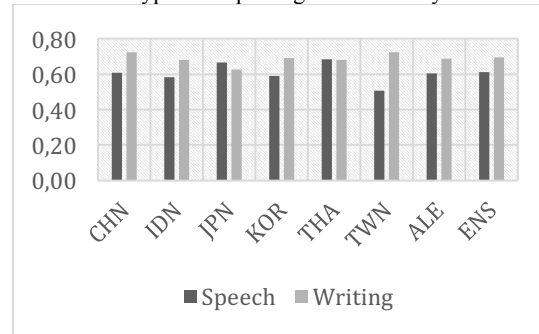

Fig. 8: Degree of variety in reporting verbs (Herdan's C) used by learners and native speakers

First, native speakers use more reporting verbs in writing than in speech. L1 speech is characterized by less varied or more phraseological use of VTHAT.

Second, learners also use more reporting verbs in writing than in speech; the degree of variety in reporting verbs is almost the same as that for native speakers as shown in Fig. 8. Among six EFL learner groups, Japanese and Thai learners use relatively more varied verbs than others do in speech, while Chinese and Taiwanese learners use more varied verbs in writing. Here we examine a speech sample by Thai learners.

> Okay, I really *agree that* – um –college student need a part-time job... because – um – it *mean that* they use – uh –they manage their time wisely, ... moreover they – um –don't need to ask parents for money because they can earn money by themselves, so they *know that* – um – uh –money is hard to find so – um –they appreciate the value of money. Um, but one thing that's – uh – more important than – um – than – then *they know that* they earn money by themselves is that – um – the experience and it does help when you do your – um – your future career such – um... (S_THA_PTJ_043_B1_2)

The speaker uses three kinds of VTHAT in his or her speech. An important fact here is that VTHAT function as a kind of glue connecting speech fragments and fillers. Without VTHAT, the speaker could not have made his or her speech, filled with speech fragments and meaningless fillers, sound cohesive and understandable. In this sense, VTHAT contributes to maintaining consistency of learners' speech.

4.3.2 Frequent reporting verbs

The top five reporting verbs in speech and writing by learners and native speakers are shown below. The codes after the verbs represent their semantic types (M: mental, S: speech act, C: communication), and the last rows represent the number of each of the three semantic types in the top five verbs.

CHN_S	IDN_S	JPN_S	KOR_S	THA_S	TWN_S	ALE_S	ENS_S
think ^{M}	agree ^{S}	think ^{M}	think ^{M}	think ^{M}	think ^{M}	think ^{M}	think ^{M}
agree ^{S}	think ^{M}	agree ^{S}	agree ^{S}	agree ^{S}	agree ^{S}	agree ^{S}	believe ^{M}
know ^{M}	know ^{M}	say ^{S}	disagree *	disagree *	know ^{M}	know ^{M}	agree ^{S}
say ^{S}	means ^{M}	disagree *	knows ^{M}	say ^{S}	hope ^{M}	say ^{S}	feel ^{M}
believe ^{M}	mean ^{M}	believe ^{M}	say ^{S}	know ^{M}	said ^{S}	believe ^{M}	know ^{M}
M 3/ S 2	M 4/ S 1	M 2/ S 2	M 2/ S 2	M 2/ S 2	M 3/ S 2	M 3/ S 2	M 4/ S 1

Table 6: Most frequent reporting verbs in learners and native speakers' speech

CHN_W	IDN_W	JPN_W	KOR_W	THA_W	TWN_W	ALE_W	ENS_W
think ^{M}	know ^{M}	think ^{M}	think ^{M}	think ^{M}	agree ^{S}	think ^{M}	think ^{M}
believe ^{M}	agree ^{S}	agree ^{M}	agree ^{S}	agree ^{S}	think ^{M}	agree ^{S}	believe ^{M}
say ^{S}	think ^{M}	say ^{S}	believe ^{M}	know ^{M}	knows ^{M}	know ^{M}	feel ^{M}
agree ^{S}	believe ^{M}	means ^{M}	know ^{M}	knows ^{M}	believe ^{M}	believe ^{M}	agree ^{S}
know ^{M}	said ^{S}	believe ^{M}	say ^{S}	believe ^{M}	means ^{M}	say ^{S}	know ^{M}
M 3/ S 2	M 3/ S 2	M 4/ S 1	M 3/ S 2	M 4/ S 1	M 4/ S 1	M 3/ S 2	M 4/ S 1

Table 7: Most frequent reporting verbs in writing by learners and native speakers

Previous studies have suggested that seven mental verbs (think, know, see, find, believe, feel, and guess), one speech act verb (say), and two communication verbs (show, and suggest) are used most frequently in L1 English in general, and that conversations and fiction are characterized by frequent use of "say," while news and academic prose are characterized by "say" and "show" respectively. In addition, learners are said to frequently use "think," "know," "say," and "hope." Our data corroborates many of these findings, but also reveals several new ones.

First, native speakers use the same five reporting verbs, four of which are mental verbs, in speech and writing most frequently. The orders are also largely the same. Although Biber et al. (1999) present different sets of reporting verbs for four text genres, our data suggests that native speakers' usage of VTHAT may be more stable or phraseological than understood before. Furthermore, native speakers are characterized by frequent use of "feel."

> ... Many *feel that* banning smoking is an attack on personal freedoms... I *feel that* banning smoking in restaurants will be good for people's health, and this in itself should be sufficient. Secondly, and I think more importantly, a ban on smoking shows to all that we have a government that cares and functions properly and intelligently...   (W_ENS_SMK_107_XX_0)

> I don't *think that* anybody should be allowed to smoke in the public place and restaurants are public place, so there should be a full smoking ban.... Personally, when I go into a restaurant and I smell cigarette, uh, I walk out and I don't *think that* it's fair that I should have to walk out because I *feel that* I should have the right to go into any restaurant without fearing for my own health and I should be able to take my young daughter as well....   (S_ENS_SMK_131_XX_0)

This shows that native speakers use not only "think" but also "feel" in an effective combination to make their claims sound softer in a pragmatically and stylistically appropriate way, while learners tend to depend solely on "think" as a softening hedge.

Second, learners use a similar set of verbs with native speakers in speech and writing. However, while native speakers are characterized by use of "feel," learners are characterized by frequent use of "say."

> ... *They say that* smoking help them many ways. *Some say that* they fell happier by smoking and forget troubles by smoking....   (W_CHN_SMK_026_A2_0)
> ... *Someone say that* part time jobs may lead the college students to a wrong way... *I have to say that* more and more students now don't know how to be with others....   (W_CHN_PTJ_328_B1_2)

The reporting verb "say" is often used in phraseological expressions such as "X say that," "I have to say that," and "it is said that." By adopting these forms, learners try to avoid taking responsibility for their own claims.

Another feature to be noted is that inflectional forms of major reporting verbs (means, knows, said) are frequently used by learners. While native speakers often tend to use VTHAT as fixed phraseological phrases, learners seem to use them in more varied ways. In addition, among six EFL learner groups, Taiwanese learners are characterized by use of "hope," which is regarded as a feature for learners in general (Biber and Reppen, 1998). Learners in Japan, Korea, and Thailand are characterized by frequent use of ungrammatical "*disagree that S+V" forms, which they seem to invent on an ad-hoc basis by overgeneralizing acceptable "agree that S+V" forms. Interestingly, this misuse is not observed in their writing, where they have enough time to monitor their L2 outputs carefully.

4.3.3 Relation between learners/ native speakers and use of major reporting verbs

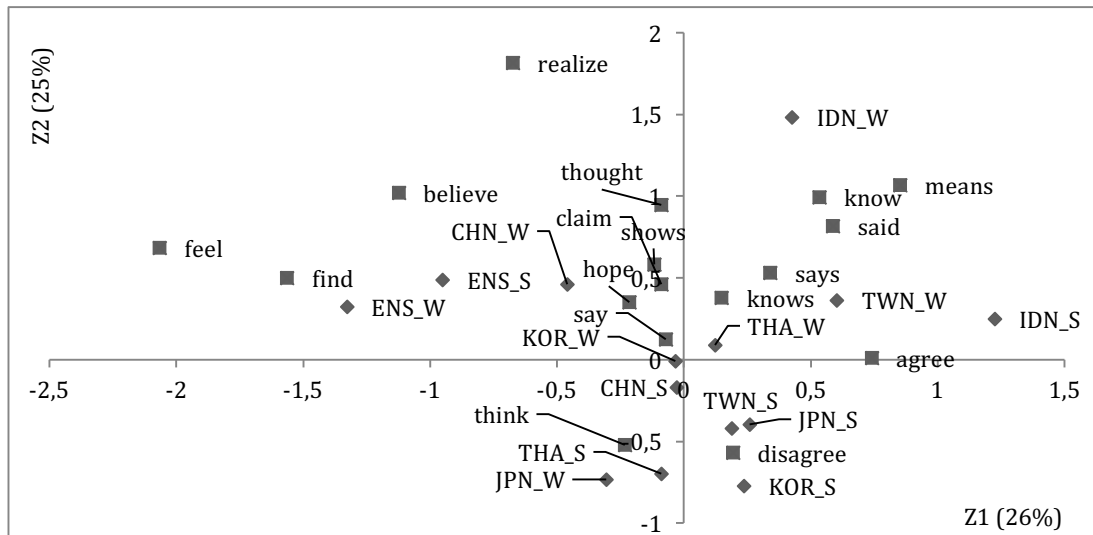The scatter plot based on the correspondence analysis is shown below.

Fig. 9 Scatter plot based on correspondence analysis

Z1 horizon axis, which explains 26% of variance, and Z2 axis, which explains 25%, distinguish all the item-category data in four quadrants. KOR_W and CHN_S seem to be located between two adjacent quadrants.

Table 8 Text clusters and characteristic reporting verbs

Quadrant 1 (Z1+, Z2+)	Quadrant 2 (Z1- , Z2+)	Quadrant 3 (Z1-, Z2-)	Quadrant 4 (Z1+, Z2-)
IDN_S/W TWN_W THA_W	ENS_S/W CHN_W	THA_S JPN_W	JPN_S KOR_S
	KOR_W	CHN_S	
means, know/ knows, says/ said	feel, find, believe, realize, etc.	think	*disagree

Our statistical analysis has illuminated several tendencies of learners and native speakers in the use of VTHAT. First, speech and writing by native speakers is clustered into the same single quadrant, which proves that native speakers use VTHAT largely similarly in their speech and writing. Further, the data has proven that they are characterized by frequent use of reporting verbs to report facts objectively (find, realize) and to soften or strengthen reporters' claims (believe, feel). By using VTHAT, native speakers seem to control speakers'/ writers' stances more carefully than learners.

Second, learners' speech and writing is often clustered into different quadrants, excluding those of Indonesian learners. This suggests that, unlike native speakers, most learners use VTHAT somewhat differently in their speech and writing. Writing by learners in Indonesia, Taiwan, and Thailand is characterized by frequent use of inflectional use of major reporting verbs (means, knows, says, and said) and speech by learners in Japan and Korea by ungrammatical use of the "*disagree that S+V" forms mentioned above.

I disagree that col....college student should have a part-time job....   (S_KOR_PTJ_096_B2_0)
I disagree that our university student take to part-time job for several reasons....
(S_KOR_PTJ_051_B1_2)
I disagree that university students have part-time job. I now, I don't have part-time job....

(S_JPN_PTJ_096_B1_1)

In addition, speech by Thai learners, as well as writing by Japanese learners is characterized by frequent use of "think."

Okay in this topic your -- you tell me that you ban smoking 100% in my country *I think that* -- *I think that* I disagree with this topic because it is for a long time that people smoking for with you are strict and *I think that* if you ban 100 smoking in the country the foreigner or the tourist from other country to go in my country. It will reduce [***] tourism because the foreigner is smoking too and *I think that* the fact –   (S_THA_SMK_037_B1_2)

While "I think that" in writing often functions as a hedge, "I think that" in speech is used to connect spoken fragments into a coherent utterance and to make time to think about what should be said next. Speakers are usually required to continue talking without stops, but this is too challenging for novice learners. Thus, some repeat "I think that" as a banal but familiar discourse device and try to make extra time. As shown in the sample above, it is sometimes repeated several times in a sequence ("I think that – I think that...").

5. Conclusions

In the current study, we reexamined the usage of VTHAT in speech and writing while focusing on the overall frequency, *that*-omission rate, and types of major reporting verbs discussed in the previous studies. Our analyses based on comparison between topic-controlled speech and writing by six Asian EFL learner groups at different L2 proficiency levels as well as English native speakers have revealed many new findings about their VTHAT use.

First, concerning RQ1 (overall frequency), we have revealed that (i) native speakers use VTHAT almost equally often in speech and writing; (ii) learners use them less often than native speakers, and (iii) at least in writing, advanced learners tend to use VTHAT less.

Second, concerning RQ2 (*that*-omission), we have shown that (i) native speakers omit complementizers much more often in speech than in writing; (ii) learners omit them more often than native speakers, especially in writing; and (iii) advanced learners tend to omit them more often.

Third, concerning RQ3 (types of reporting verbs), we have confirmed that (i) the degrees of variety in reporting verbs are almost the same for learners and native speakers; (ii) the top frequent verbs (think, believe, know, and agree) are common for learners and native speakers in speech and writing; (iii) native speakers' outputs are characterized by the use of a common set of verbs for speech and writing and frequent use of reporting verbs presenting facts objectively (find, realize) or adjusting stances (believe, feel), while learners' outputs are characterized by different sets of verbs for speech and writing, inappropriate "*disagree that" forms, and inflectional forms of basic verbs.

VTHAT are one of the most common English constructions widely seen across varied genres, but their discourse functions and formal features have not been wholly elucidated yet and therefore they are not necessarily presented appropriately in TESOL. The findings from the current study based on a reliable international learner corpus will help EFL teachers and learners to deepen their understanding about the basic questions concerning major structural types, common controlling verbs, and genre dependency of VTHAT, which should also lead to sophistication in the description of grammar and reference books.

References

Aijmer K. (2001). I think as a marker of discourse style in argumentative Swedish student writing. In K. Aijmer (Ed.), *A wealth of English* (pp. 247-257). Goteborg: Acta Universitatis Gothobrugensis.

Biber (2006) *University language: A corpus-based study of spoken and written registers*. Amsterdam: John Benjamins.

Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge, UK: CUP.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow: Pearson Education.

Biber, D., & Reppen, R. (1998). Comparing native and learner perspectives on English grammar: A study of complement clauses. In S. Granger (Ed.), *Learner English on computers* (pp. 145-158). London: Longman.

Carter. R., & McCarthy, M. (2006) *Cambridge grammar of English: A comprehensive guide*. Cambridge: Cambridge University Press.

Chen, M-Y. (2015). Spoken discourse markers in English writing of Taiwanese students. *Journal of Modern Education Review*, 5(5), 453–460.

Granger, S. (1998). Prefabricated patterns in advanced EFL advanced writing: Collocations and formulae. In A. P. Cowie (Ed.), *Phraseology* (pp. 145-160). Oxford: Oxford University Press.

Granger S., & Tyson, S. (1996). Connector usage in the English essay writing of native and non-native EFL speakers of English. *World English, 15*, 19–29.

Hinkel, E. (1999). Objectivity and credibility in L1 and L2 academic writing. In E. Hinkel (Ed.), *Culture in second language teaching and learning* (pp. 90-108). Cambridge: Cambridge University Press.

Ishikawa, S. (2013). The ICNALE and sophisticated contrastive interlanguage analysis of Asian Learners of English. In S. Ishikawa (Ed.), *Learner corpus studies in Asia and the world, 1* (pp. 91-118). Kobe, Japan: Kobe University.

Ishikawa, S. (2014). Design of the ICNALE-spoken: A new database for multi-modal contrastive interlanguage analysis. In S. Ishikawa (Ed.), *Learner corpus studies in Asia and the world, 2* (pp. 63-76). Kobe, Japan: Kobe University.

József, H. (2001). *Advanced writing in English as a foreign language: A corpus-based study of processes and products*. Pécs: Lingua Franca Csoport.

Natsukari, S. (2012). Use of I in essays by Japanese EFL learners. *JALT Journal, 34*(1), 61-78.

Sakaue, T. (2013). Nihonjin eigo gakushusha no essei ni mirareru kyoki hyogen no bunseki. *Hiroshima Gaikokugo Kenkyu, 16*, 159-169. [An analysis of collocations in intermediate-level EFL learners' writing]

Tono, Y. (2004). The NICT JLE corpus ni miru eigo gakushusha no happyo goi no siyo jyokyo. [Use of productive vocabulary seen in the NICT JLE corpus]. In E. Izumi, H. Isahara, K. Uchimoto (Eds.), *Nihonjin 1,200-nin no eigo supikingu kopasu* (pp. 96-112). Tokyo, Japan: Arc. [Speaking corpus collecting 1,200 speeches by Japanese learners]