

## **Corpus–Based Comparative Analysis of Reporting Verbs in English Essays Written by Four Language Nationalities**

**YAMASHITA, Miho**  
Tezukayama University

### **Abstract**

Learner corpus study has been gaining popularity in recent years both internationally and domestically, revealing learner characteristics of both speaking and writing. This study examines the reporting verbs used to express opinions in argumentative essays written by college students with four language backgrounds. Relative frequencies of reporting verbs distinctive to each nationality are investigated both quantitatively and qualitatively. Significant and different features of types and usage of verbs are revealed, especially in Japanese learners' writing; Japanese learners of English seem to have limited knowledge of academic vocabulary. They particularly use the word "think" to express their opinions. Native speakers of English use more varieties of verbs appropriately in their statements. The findings of this study suggest several ideas or approaches for teaching Japanese learners of English how to write English essays and further research implications of this type of comparative learner corpora analysis are pointed out.

**Key words:** learner corpora, quantification of contingency table (QCT), log-likelihood ratio statistics, concordance.

### **Introduction**

Corpus Linguistics is a rather new field of discipline that emerged in the 1970s. "Learner corpora" started to become the focus of attention among researchers as the huge influx of international students in English-speaking countries necessitated researchers to identify the learners' spoken and written features. Compilations of major learner corpora started around the world and their analyses have followed. Major findings from previous studies have revealed typical learner texts' features, which are quite different from those found in the texts created by native speakers of English.

The prime features of learners' written texts include simple syntactic structures with a limited lexical repertoire (Carlson, 1988; Leki, 1991; Hasselgren, 1994; Ringbom, 1998, Granger and Rayson, 1998; Read, 2000; Hinkel, 2003). Ringbom (1998) examined the overuse and underuse of linguistic features in *the International Corpus of Learner English* (ICLE hereafter), arguing that these features are attributable to "L1 transfer" where learners try to find the English words corresponding to those of their L1. When they cannot find the specific word to refer to, they tend to underuse or avoid using the particular English word or they depend on the words which are familiar. The author mentioned that because of limited lexis,

the general impression of learners' texts in comparison with those written by native speakers of English is "dull, repetitive and unimaginative." The other key feature of learners' texts is confusion of registers (Granger and Rayson, 1998; Lorenz, 1999; Aijmer, 2002; Ishikawa, 2008). Learners tend to overuse words and phrases which are more likely to appear in speech, and underuse more formal expressions typical of academic writing. The tone of speech in the writing is particularly identifiable among European nationalities (Granger and Rayson, 1998; Lorenz, 1999), and can be ascribed to their educational backgrounds, which generally tend to focus on conversational English, and to similar etymology. As for another discourse feature of overly personal involvement in the written texts, Petch-Tyson (1998) has revealed a higher use of indicator of personal involvement such as first person singular *I* across European nationalities than the U.S. The essays written by Japanese university students show a similar feature of high visibility of writer (McCrostie, 2008). Their texts particularly show lack of hedging, which can soften the tone of strong personal opinion often used by native speakers of English (Hyland, 2005; Kobayashi, 2009). These previous studies have revealed typical features in learners' texts. However, the studies of comparative corpus analysis across different nationalities including Asian nationalities and locating the Japanese learners of English properly in the international context are scarce.

This study focuses on the reporting verbs used by EFL students. The appropriate use of reporting verbs is crucial in establishing the writer's own claims (Bloch, 2010). However, learners of English often find it difficult to choose among the wide variety of reporting verbs (Hyland, 2005), and this happens even among advanced English learners (Bloch, 2010). Most of the previous research explored the usage of reporting verbs in citation of academic papers (Thompson and Ye, 1991; Hyland, 1999; Charles, 2006; Bloch, 2010). The authors created a corpus of research articles, from which they identified the kinds of reporting verbs used in citations and categorized them for future use by EFL students. Hyland (1999) investigated 80 research articles from different disciplines and found clear disciplinary differences, especially between hard and soft disciplines. Charles (2006) explored the phraseological patterns used in reporting clauses in two disciplines, according to semantic groups categorized in Collins COBUILD grammar patterns 1: Verbs (1996). These are all investigations into academic papers or theses; however, investigation into reporting verbs used in argumentative essays, which are one of the most frequently assigned tasks in EFL writing classes, has been scarce. An exception is the study by Neff et al. (2003), in which they investigated nine reporting verbs used in argumentative essays written by Dutch, French, German, Italian and Spanish learners of EFL, compared with American. The study found characteristic usages (overuse and underuse) in each European nationality and EFL students' general lack of knowledge for the different evaluative meanings attached to different reporting verbs.

The purpose of the present study is to explore the different usage of reporting verbs used in argumentative essays among four different groups: native speakers of English (American and British) and Japanese, Chinese and German learners of English. With both quantitative and qualitative analyses, this study ultimately seeks to clarify the usage typical to Japanese by comparing different nationalities, and highlight the necessity of teaching the appropriate use of reporting verbs. Thus, the following two research questions guided the entire analysis:

1. What types of reporting verbs are unique in learners' essays, especially in the Japanese

essays?

2. How different are they from those used by native speakers of English?

## Methodology

### Corpus Data and Data Description of Each Nationality

In this research, two types of corpora were used. The first is ICLE (*International Corpus of Learner English*) version 2, which was compiled by the members of *Center for English Corpus Linguistics (CECL)*. Included in the corpus are the essays written by 16 international college students with upper levels of English. Students are EFL learners rather than ESL. Texts used in this study are all argumentative essays on various types of topics. The average length of the texts is about 617 words and the number of words from each language nationality is approximately 200,000.

In order to identify the distinctiveness of essays by Japanese (NJS hereafter), the data of both Chinese (NCS hereafter) and Germans (NGS hereafter) were selected as comparative data. The reasons why the data of these two nationalities were picked are twofold: one is that Chinese is the only Asian data available to be compared with the Japanese data and the other is that German is the language most similar to English in terms of grammatical and lexical structures. The other corpus, used as a reference, is LOCNESS (*Louvain Corpus of Native English Essays*), also compiled by the group mentioned above. This is a corpus of argumentative essays written by native English speakers (NES hereafter): British and American college students. This corpus is combined as one corpus of native speakers of English in this study. The composition of the data is shown in Table 1.

Table 1

*Composition of the Number of Essays and the Number of Words by Four Nationalities*

	LOCNESS	Japanese	Chinese	German
Number of Essays	436	366	982	437
Number of Words	324,304	198,241	490,617	229,698

### Selection of Verbs to Investigate

In order to determine the reporting verbs to be investigated, Hinkel (2002)'s study was replicated. She selected indirect verbs that follow *that*-clauses as objects. These are the verbs which belong to three semantic categories: factual verbs as public and private, and suasive verbs (Quirk et al., 1985).

Public verbs refer to actions that can be observed publicly and that are used to introduce reported statements. The verbs chosen for this study are as follows:

*acknowledge, add, admit, affirm, agree, allege, announce, argue, assert, bet, boast, certify, claim, comment, complain, concede, confess, confide, confirm, contend, convey, declare, deny, disclose, exclaim, explain, forecast, foretell, guarantee, insist, maintain, mention, object, predict, proclaim, promise, pronounce, prophesy, protest, remark, repeat, reply, report, retort, say, state, submit, suggest, swear, tell, testify, vow, warn, and write*

Private verbs express intellectual states and non-observable intellectual acts that are “private.” The private verbs selected for the present study include:

*accept, anticipate, ascertain, assume, believe, calculate, check, conclude, conjecture, consider, decide, deduce, deem, demonstrate, determine, discern, discover, doubt, dream, ensure, establish, estimate, expect, fancy, fear, feel, find, foresee, forget, gather, guess, hear, hold, hope, imagine, imply, indicate, infer, insure, judge, know, learn, mean, note, notice, observe, perceive, presume, presuppose, pretend, prove, realize, reason, recall, reckon, recognize, reflect, remember, reveal, see, sense, show, signify, suppose, suspect, think, and understand.*

Suasive verbs function as causal or “mandative” (Quirk et al., 1985, p. 1182) and express a directive to or intention for change. The suasive verbs chosen for the present study include:

*agree, allow, arrange, ask, beg, commend, concede, decide, decree, demand, desire, determine, enjoin, insure, entreat, grant, insist, instruct, intend, move, ordain, order, pledge, pray, prefer, pronounce, propose, recommend, request, require, resolve, rule, stipulate, suggest, urge, and vote*

Some of the verbs overlap and are counted as one word. These three groups are combined into the final list of 150 reporting verbs.<sup>1</sup>

### Procedure of Data Processing and Analysis

This section describes how the data was processed for the present analysis. After some unnecessary codes originally inserted in the texts such as <\*>, <quote> for quote, and <R> for bibliographic references were eliminated, all the texts were tagged by Go-tagger (ver 0.7) created by K. Goto (2006) in order to pick up just the verbs. The tagged data were then all processed by KWIC Concordance for Windows (ver 4.7), a software developed by S. Tsukamoto (2006) to create an alphabetical frequency list of all the verbs. In the present research, this software was used for processing the data and also creating the concordance lines showing the contexts in which certain words occur. Then, the verbs were extracted from the alphabetical frequency list of each corpus, and all the frequencies of the inflected variants were added together manually for each corpus. Finally, a cross table indicating how each reporting verb is used across four nationalities was created.<sup>2</sup> Table 2 is a part of this cross table showing only the first and the last 10 items due to the limited space allowed.

Table 2

*A Part of the Cross Table Showing Frequencies of Reporting Verbs across 4 Nationalities*

Reporting Verb	Subclass	Total	NES	NCS	NGS	NJS
NES		16	16	0	0	0
NCS		6	0	6	0	0
NGS		7	0	0	7	0
accept	pv	325	179	78	34	34
acknowledge	pb	24	13	4	4	3
add	pb	89	37	35	0	17
admit	pb	136	40	40	33	23
affirm	pb	7	4	2	0	1
agree	pb&sv	162	80	1	34	47
allow	sv	562	265	203	52	42
announce	pb	29	11	7	6	5
anticipate	pv	6	4	1	0	1
argue	pb	545	165	339	21	20
suspect	pv	21	6	7	4	4
swear	pb	7	1	1	4	1
tell	pb&sv	546	148	87	162	149
testify	pb	4	3	0	1	0
think	pv	3149	430	901	306	1512
understand	pv	479	142	63	39	235
urge	sv	42	3	36	1	2
vote	sv	32	17	1	6	8
warn	pb	24	5	16	3	0
write	pb	423	128	44	46	205
Total		6641	1697	1872	763	2309

In order to see the general tendencies and characteristics of the four nationalities, all the verbs in the cross table were processed by means of Quantification of Contingency Table (QCT hereafter), one of the functions prepared in the suite of programs called *FAT (Freq Analysis Toolkit)* developed by Nakamura (2006). QCT is a kind of technique of multivariate analysis used to examine the relationships among texts (i.e., texts of the four different nationalities in the present study), the relationships among linguistic features (i.e., reporting verbs), and the relationship between texts and linguistic features, which cannot be identified only through the reading of texts (Moriwaki, 2011). The gist of this complicated statistical procedure is to give several sets (called axes) of category weights (quantities given to texts in the present case) and sample scores (quantities given to linguistic features) simultaneously so that the first set may produce the highest correlation coefficient between category weights and sample scores and the second set the next highest correlation coefficients and so on. Texts given close quantities are considered to be qualitatively similar and those given distant

quantities are considered to be qualitatively different. In the same way, verbs given close values are also considered to be qualitatively similar and those given distant values are considered to be qualitatively different. The first two or three sets thus obtained, depending upon the proportions accounted for, are usually used for analysis through plotting the quantities on two- or three-dimensional space as shown below (Figures 1, 2, 3 and 4).

Although QCT displays the relationships among texts and the reporting verbs visually, it does not provide rigorous statistical measures to determine whether or not the facts obtained for the verbs in question are statistically significant. Therefore, Log-Likelihood Ratio Statistics<sup>3</sup>, a statistical measure often used to pick up the key words in a particular corpus in relation to a reference corpus, was later used to pick up characteristic reporting verbs for each nationality in comparison with native speakers of English. Finally, KWIC Concordance for Windows and AntConc ver 3.2.1 (A. Laurence, 2000) were used in order to examine how each characteristic verb is used in concrete contexts. To compare the usage of reporting verbs shown in the high school texts, n-grams were also examined.

## Results and Discussion

The corpus analysis of the selected reporting verbs is presented in this chapter, and the research questions posed in the previous chapter are addressed. Firstly, quantitative results obtained from QCT are examined, followed by a statistical analysis using log-likelihood ratio. Secondly, qualitative results are discussed in terms of concordance and collocation of the particular verbs obtained from a quantitative analysis.

### The Results of QCT

#### Category Distribution

Figure 1 shows the distribution of the four nationalities based upon the relative frequencies of reporting verbs they used. Each nationality is distinctively separate, meaning that there are verbs dominantly used by each nationality. On Axis 1, NJS and NCS are completely in the opposite range with NES in the center. NJS is within the negative range while NCS is within the positive range. This means that the verbs used by these three nationalities are quite different from one another. On Axis 2, NJS and NCS are within the same positive side while NES and NGS are positioned within the negative side. Particularly compared to NES, NJS is situated almost in opposite sides along the second axis, indicating that NES and NJS use verbs in completely different ways. In terms of the size of the circles which shows the volume of the verbs used, the relative volume of the verbs used by NGS is far smaller than the rest of three, reflecting the total number of tokens in the corpus as shown in Table 1.

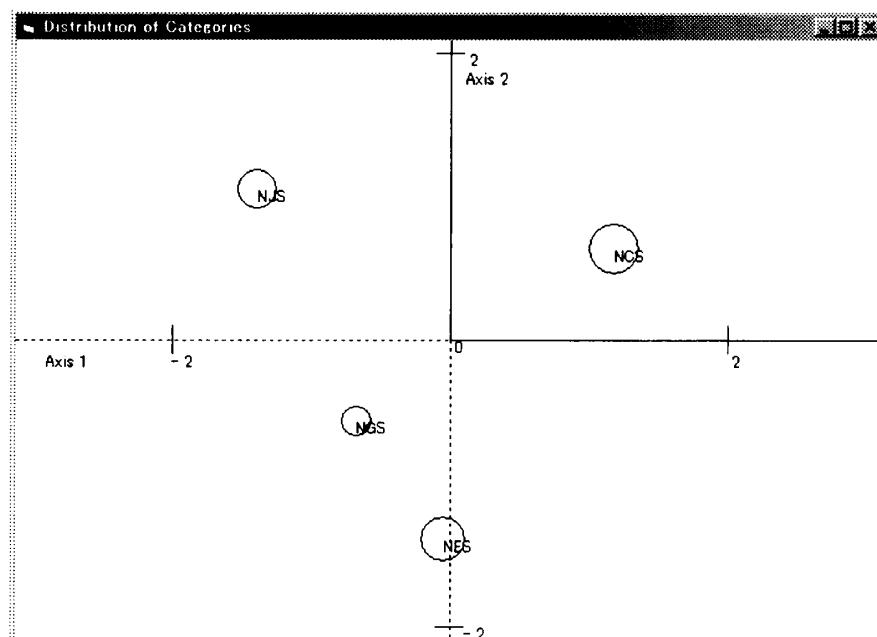


Figure 1. Result of QCT: Distribution of Four Nationalities

Although Figure 1 shows that the way in which different nationalities use verbs is different, how or why they are different is not really explicated. The reason why these four nationalities are distributed in this way is explained by looking at the sample distribution, i.e., the distribution of verbs based upon the frequencies of verbs used by different nationalities. This will be examined in the following section.

### Sample Distribution

The sample distribution in Figure 2 and its extracts Figures 3 and 4 show the relative frequencies of the verbs and how they are used by the four language nationalities as reflected in their positions. In these figures, the verbs close to the origin indicate that they are evenly used by the four nationalities. The verb *mean*, for example, is close to the origin, indicating that this is more or less equally used by the four language nationalities. Other verbs dispersed away from the origin reflect the way in which they are used characteristically by each nationality. For example, *think* is positioned closer to or corresponding to the area of NJS in Figure 2 and 3 with no other words around it, meaning that it is dominantly used by Japanese. In the area around NES depicted in Figure 4, *pray*, *testify*, *insure*, *exclaim*, *disclose*, *perceive* and so on can be seen. Those verbs are the ones mainly used by NES and rarely used by other nationalities. In the same way, how each verb is used by four nationalities can be grasped by its relative location in reference to category distribution in Figure 1.

NII-Electronic Library Service



QCT, thus, turned out to be quite effective in visually grasping how each nationality used these reporting verbs, but it is only a rough picture based upon the relative frequencies across four nationalities. To clarify the verbs distinctively used by each nationality in a more rigorous statistical way, particular verbs are picked up by means of log-likelihood statistics processing the frequencies of each of the L2 learners (NJS, NCS, or NGS) against those of NES. These verbs will be discussed in the following sections.

QCT identifies *think* as a verb unique to Japanese. According to Figure 3, it is the only frequent verb located closer to NJS than other nationalities. Although this verb is also used by other nationalities (901 times by NCS and 306 times by NGS), it stands out in Japanese essays when its frequency, 1512, is compared with 901 of NES. Table 3 shows the result of LL ratio statistics which is sorted according to the signed LL, so that the verbs distinctively used by NJS are displayed with the highest statistical significance from the top of the list. The verb *think* is the most significant verb overused by NJS ( $p < .0001$ ) and it is followed by *learn*, *understand*, *know*, *write*, *say*, *check*, *recommend*, *forget*, and *hear*. These verbs stand out in Japanese essays when compared with NES, but they are also the ones that stand out in other nationalities although they are not shown on the top part of the list; therefore, in this section *think* is examined more closely.

Table 3

*Frequencies of Verbs Characteristically Used by NJS against NES and their LL Ratio Statistics*

Reporting Verb	NJS	NES	Total	Signed LL	Probability
think	1512	430	1942	1081.4960	3.5E-237
learn	542	127	669	407.8895	1.06E-90
understand	235	142	377	53.0958	3.18E-13
know	444	355	799	46.4401	9.45E-12
write	205	128	333	42.8843	5.81E-11
say	566	495	1061	41.2568	1.33E-10
check	44	9	53	34.6885	3.87E-09
recommend	24	1	25	32.3272	1.30E-08
forget	67	31	98	24.0099	9.58E-07
hear	130	88	218	22.0652	2.64E-06
notice	35	17	52	11.6011	0.000659
repeat	21	8	29	9.6847	0.001858
insist	28	16	44	6.9408	0.008425
certify	7	1	8	6.6576	0.009873
state	39	128	167	-30.9966	2.58E-08
determine	9	64	73	-34.4370	4.40E-09
deny	5	53	58	-35.5557	2.48E-09
reveal	4	54	58	-40.0326	2.50E-10
feel	183	408	591	-43.2991	4.70E-11
claim	4	67	71	-53.4707	2.62E-13
prove	16	118	134	-65.2715	6.53E-16
discover	0	62	62	-71.8863	2.28E-17
realize	48	210	258	-75.1443	4.38E-18
accept	34	179	213	-76.9137	1.79E-18
believe	118	368	486	-83.6666	5.86E-20
argue	20	165	185	-98.0291	4.12E-23
show	84	349	433	-119.0760	1.01E-27
allow	42	265	307	-132.9790	9.14E-31
see	165	633	798	-200.5800	1.56E-45

In order to analyze the contexts in which the verb *think* is used by Japanese writers, the words collocated with *think* were obtained using AntConc. AntConc uses MI-score for statistical validation for collocates with a span of 4 words before the node word. The result shows that the first person singular pronoun “I” (occurring 1067 times) is highly collocated with “think,” followed by “we” (89), “people” (63), “you” (40), “they” (37), “Japanese” (30), “students” (19), and “he” (12). This means that Japanese learners use “I” far more frequently than any other words like “we,” “people,” “you,” or “they,” etc.

Concordance lines allow patterns to be seen in the contexts when particular words are used. The most popular expression is ‘I think S (any subject) *should* ... .’ One example of this in Figure 5 is ‘I think English *should* be set as the second official language.’ The modal

auxiliary *should* is frequently associated with this pattern. Japanese learners of English are considered to use this expression to state a strong argument in their essays. Considering this verb shows the highest number in frequency and pattern, it is believed that *think* is used by Japanese learners almost all the time to voice their personal opinion in essays.

Another popular expression is ‘I think that *I want to* ... .’ Interestingly enough, this expression is one that is quite peculiar to NJS. Not a single example of this expression has been found in the NES essays: Native speakers of English just use ‘I want to ... .’ It seems that Japanese learners use “I think” as a kind of a “hedge” to soften their statements (Kobayashi, 2009), eventually making them sound overly redundant. Two other expressions to be noted are ‘I think it is ADJ + to/that ... .’ and ‘I think that *there is/are* ... .’ Favorite adjectives used in this pattern as shown in Figure 6 include; *important*, *necessary*, *good*, *bad*, *right*, or *difficult*. Comparative and superlative forms are also used as in “I think that it is *more important* to master English.”

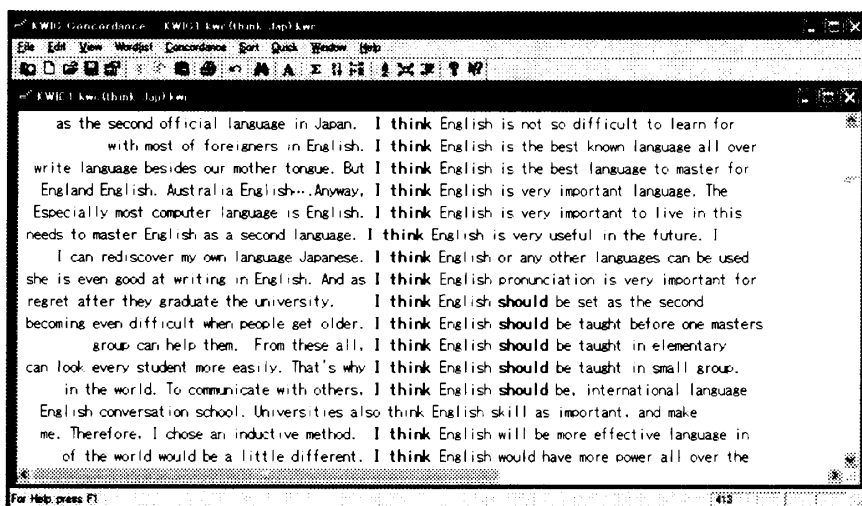


Figure 5. KWIC Concordance of ‘think | thinks | thought | thinking’ from Japanese Essays (I think ... should ...)

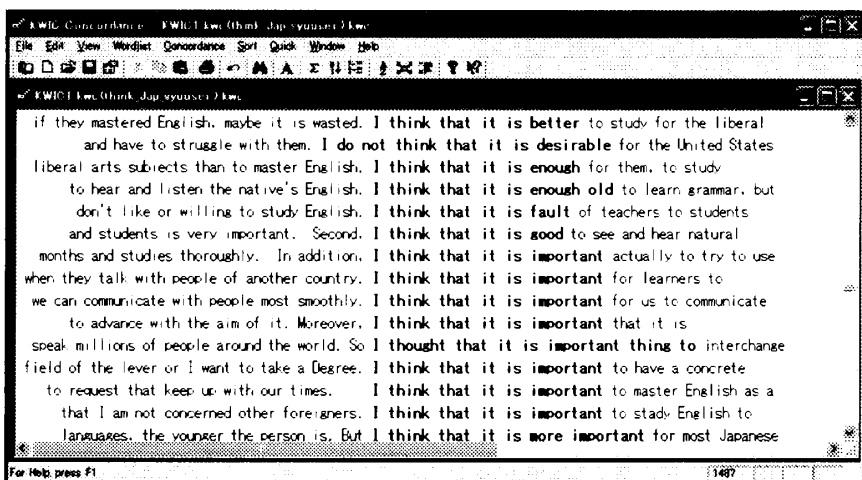


Figure 6. KWIC Concordance of ‘think | thinks | thought | thinking’ from Japanese Essays (I think it is adjective + to / that ...)

Some typical phrases reveal other fixed expressions such as 'I think that *there are* many problems / reasons' as in the sentence "I think that *there are* many reasons to support his idea." These expressions seem to be often introduced in English textbooks used in high schools in Japan. Therefore, Japanese college students may use them as a type of fixed expression in their argumentative essays. In order to justify this observation, the present author examined the textbooks in the discussion section for the findings obtained from NJS and NES.

### Distinctive Verbs Used by NES Compared with NJS

According to QCT in Figures 2 and 4, various kinds of verbs are concentrated around the area typically used by native writers of English. LL ratio statistics confirm them and they are listed in the second half of Table 3. The verbs overused by NES are the ones placed on the bottom of the list with a high negative LL value, and these are the ones underused by NJS in comparison with NES. Apparently, some of the academic words which Japanese students find difficulty in using are placed on the bottom of the list; however, among them, such verbs as *see*, *allow*, and *accept* are often used in ways other than reporting statements and are very difficult to discern. Thus, these verbs are excluded from discussion in this section.

One of the verbs most distinctively used is *argue*. AntConc shows the words typically collocated with this particular verb with a span of 4 words. They are "people" (occurring 15 times), "it" (12), "many" (11), "they" (7), "some" (6), "opponents" (5), "one" (4), "activists" (2) and "researchers" (2), and they are mostly used as subjects of this verb. It does not usually occur with "I," indicating that writers use this verb from a third person viewpoint. In addition, the subject in the sentence is often followed by modal verbs like *should* or *would*. These are used as "hedging" to tone down the writer's claim (Hyland, 2005). As indicated in Figure 7, the form '*would* + *argue* that ...' is popular. One typical example seen is "Very few people *would argue* that there is a moral problem."

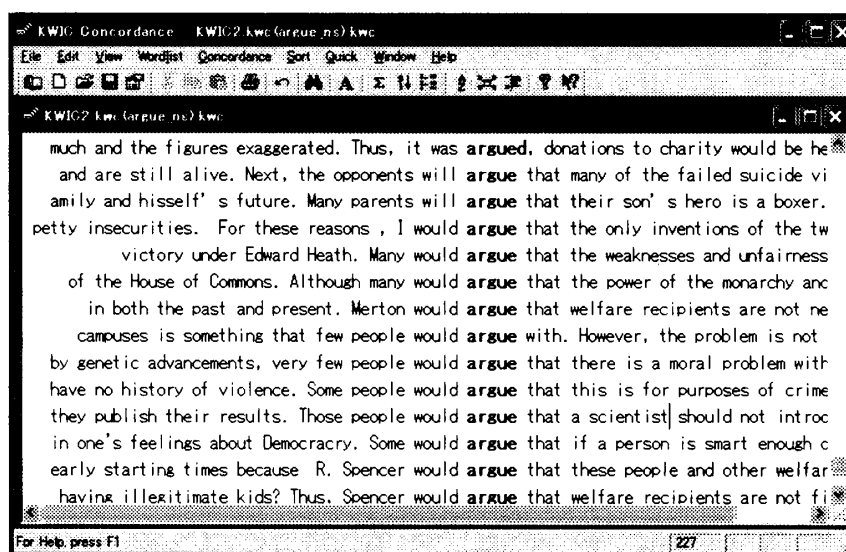


Figure 7. KWIC Concordance of 'argue | argues | argued | arguing' from NES Essays

The verb *claim* is also used for asserting one's opinion from someone else's view. Some of the collocates are "government" (2), "coalition parties" (1), "reporters" (1), "supporters" (1), "proponents" (2) and "opponents" (1), all of which imply reliable source of information to support the opinion.

Another notable verb is *believe*. Similar to *argue*, this particular verb *believe* has a strong meaning when expressing one's ideas. Native speakers of English seem to use these verbs when they are certain that their argument is true. This verb collocates dominantly with "I" (46), followed by "they" (12), and "people" (10). Although the number is small, such words as "academy" (2), "critics" (2), "followers" (2), "opponents" (2), "researchers" (2), "others" (1), and "scientists" (1) are also used. As seen in *claim*, "academy," "researchers," or "scientists" are considered as reliable sources of information, therefore, NES essays can be considered very persuasive.

Unlike *argue* or *believe*, NES use *feel* when they simply want to convey their ideas. *Feel* collocates with "I" (66), "people" (16), "many" (10), "you" (5), "who" (4), and "we" (2). Patterns such as 'I feel that ...,' 'I feel + adjectives' and 'I feel like V-ing' form can be seen in the concordance line. It seems that native speakers of English use *feel* in a somewhat similar way to *think* when they state their idea without concrete evidence.

Other than the verbs to state one's own personal opinions, native speakers of English produce a variety of verbs appropriate for academic essays. The verbs *realize* and *discover* are predominantly used by NES, compared to NJS. Both *prove* and *show* seem to be used to provide supporting evidence for claims. Evidence offered by reliable sources such as "government," "advocates of censorship," or "research" endorse the writer's personal opinions. This type of statement is apparently missing in Japanese essays. One of the examples from the concordance line with the word *prove* is "These people argue that there is evidence to *prove* that the goal of abstinence is realistic."

### Discussions of the Findings obtained from NJS and NES

The comparison between Japanese texts and English native speakers' texts show a critical difference in terms of syntactic, lexical and rhetorical features. Overall, Japanese learners' essays show a simple syntactic structure with limited lexical variety and a speech-like nature, all of which have been identified in L2 texts (Carlson, 1988; Leki, 1991; Ringbom, 1998; Granger & Rayson, 1998; Read, 2000; Hinkel, 2003)

One significant lexical feature is that Japanese overly use the word *think* with the fixed pattern of 'I think (that)....' Considering the fact that there are no other distinct words used only by the Japanese learners as shown by QCT and LL Ratio Statistics, it seems that Japanese learners of English heavily rely on *think* to express their opinions in writing. This result is in agreement with Ishikawa (2008) which investigated written data of Japanese college students. This can be attributed to the L1 transfer (Granger, 1998b; Ringbom, 1998). Japanese college students simply translate the Japanese equivalent "*to omou*" into *I think*. This is closely related to the fact that lexical knowledge of Japanese learners of English is quite restricted. McCrostie (2008) sees a similar tendency among his Japanese university students who fail to use a variety of verbs to state their claims other than *think*. They may depend on what is familiar and stick to the words they feel safe with (Hasselgren, 1994). Even though there are

other verbs associated with the Japanese expression “*to omou*” such as *consider*, *guess*, *suspect* or *suppose*, depending on the situation used, it seems that Japanese college students do not learn them, nor can they distinguish between them or use them in the right context. This is a common characteristic identified among L2 learners of English. Granger (1998b) refers to “cover-all” *think* used frequently by L2 learners of European nationalities.

Native speakers of English, on the other hand, have a wider knowledge of academic words, which include *argue*, *claim*, *believe* and *feel*. They distinguish these words and use them appropriately. For example, *argue*, *claim* and *believe* are used to convince readers to believe in their argument, while *feel* is used to express their personal feelings or ideas. In addition, their essays appear to be logically constructed as their opinions have supporting evidence, which is presumable from the verbs *show* and *prove*. This shows a significant contrast with Japanese essays that lack these verbs critical for rhetorical organization.

The results mentioned above mean that learners’ texts create a negative impression to native speakers of English. Ringbom (1998) argued that “limited lexical variation that learners have in comparison to native speakers is a main reason for the general impression of learner language as dull, repetitive and unimaginative.” Other studies (Carlson, 1988; Leki, 1991; Read, 2000) suggest that limited lexical repertoire results in vague and less sophisticated prose relative to that of native speakers, and this also gives an impression of verbosity to the learners’ texts in that many words are unnecessary in their contexts.

Another finding about the Japanese learners’ usage of *think* is that the verb often comes with the subjects “I” or “We.” It can be explained as high writer’s visibility, meaning that Japanese essays are too personal (McCrostie, 2008). Technically, “I” and “we” are visible even in native speakers’ texts; however, sentences starting with “I” or “we” seem more conspicuous in Japanese texts. Hinkel (2002) argues that this is apparent especially in the form of ‘I think that we *should* / *must* ...’ She says this form is highly influenced by Asian culture which is recognized being as “group-oriented,” referring to the collocated modal verbs *should* and *must*. Japanese students may tend to express their opinions strongly as a member of their community.

Concordance lines revealed typical patterns used by the Japanese learners of English and they seem to use some fixed patterns. This can be attributed to English textbooks in high school. They might have been taught using certain text book forms and memorized them as fixed expressions. In order to test this assumption, 15 high school English textbooks were analyzed. These textbooks were all certified by the Ministry of Education, Culture, Sports, Science and Technology. Figure 8 is the result of n-gram analysis of 15 high school English textbooks using AntConc. N-gram shows the recurrent word combinations used in the texts by their frequency. The pattern ‘I think we should’ is displayed; however, the frequency is quite low and so does not really explain the frequent usage. Future investigation is necessary to resolve this matter.

One more prominent result is that the word or structures identified in Japanese texts are more typical of speech than of academic writing. High writer’s visibility in Japanese texts, which is observed by the frequent use of the first person singular *I*, indicates a conversational style (Petch-Tyson, 1998; McCrostie, 2008). “I think,” “I want” and “I would” found in their texts can be frequently observed in speaking (Ishikawa, 2010). This tendency is also prevalent

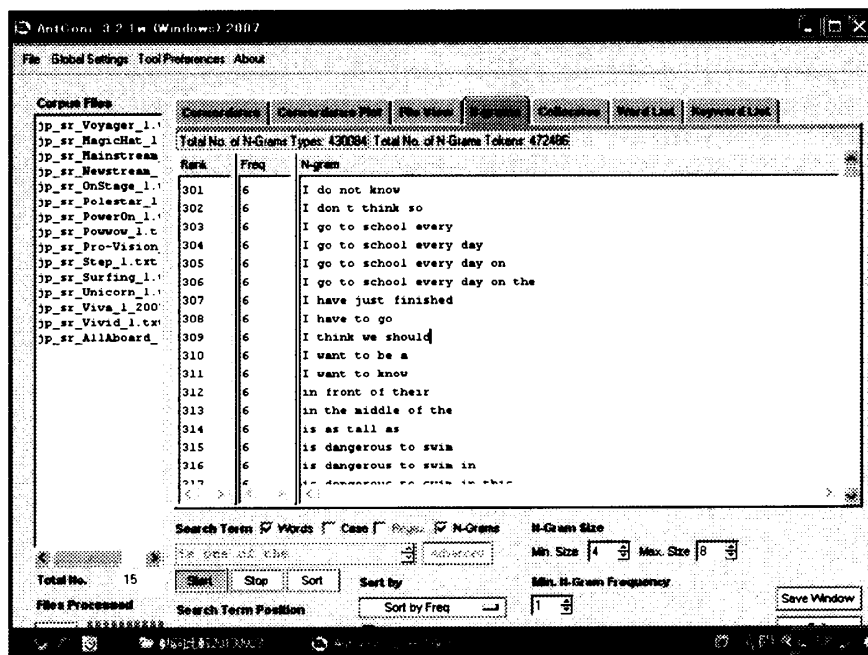


Figure 8. N-gram Analysis of 15 High School English Textbooks Using AntConc

among European nationalities and may be explained by the non-native learners' lack of intuition regarding the distinction between casual speech and formal writing (Lorenz, 1999). Japanese students might not distinguish the two registers, incorporating speech voice into academic essays. This is attributable to the influence of teaching, which has been more focused on oral skills and less focused on academic writing for the past decade of Japanese English education.

### Distinctive Verbs Used by NCS Compared with NES

The results of QCT and LL ratios revealed that Chinese writers use a variety of verbs for their arguments in comparison with native speakers of English as listed in Table 4 ( $p < .0001$ ). Among them, *bet* comes on the top of the list, but should be excluded for discussion because this verb is used to "risk a sum of money", being influenced by the title of the essay.

One of the features of Chinese usage is that they use suasive verbs such as *suggest*, *recommend*, *urge* and *propose* far more frequently than other nationalities. Chinese writers tend to propose personal ideas using, for example, *suggest* with collocates such as "I" (28), "people" (10), "they" (9), "supporters" (2), "parents" (2), "leaders" (2), "management" (1), and "companies" (1). On the other hand, they may use factual verbs such as *indicate*, *argue*, and *claim* with the third person subject to give an objective opinion to support their ideas similar to the usage in NES essays. The verb *claim*, for example, collocates with such subjects as "people" (18), "respondents" (17), "researchers" (11), "companies" (8), "someone" (2), "students" (2), "doctors" (2), "they" (2), and "I."

Another finding is a wide range of forms or fixed expressions identified for *conclude* and *mention*. Some examples for *conclude* include; '... leads me to conclude that ...' (35), 'I can conclude that ...' (15), and 'It was concluded that ...' (8), and 'It can be concluded that ...'

Table 4

*Frequencies of Verbs Characteristically Used by NCS against NES and their LLRatioStatistics*

Reporting Verb	NCS	NES	Total	Signed LL	Probability
bet	385	5	390	420.9016	1.55310E-93
indicate	202	10	212	178.2452	1.17105E-40
conclude	204	30	243	113.6192	1.57899E-26
suggest	258	52	310	112.5434	2.71652E-26
mention	143	32	175	55.91473	7.56833E-14
consider	376	173	549	42.65164	6.54095E-11
argue	339	165	504	32.12717	1.44403E-08
recommend	32	1	33	30.88755	2.73421E-08
urge	36	3	39	26.67578	2.40618E-07
propose	64	19	83	17.57200	2.76632E-05
claim	146	67	213	16.38131	5.17934E-05
hear	27	88	115	-47.9107	4.46073E-12
understand	63	142	205	-49.9182	1.60290E-12
recognize	13	74	87	-60.7546	6.46532E-15
write	44	128	172	-62.1981	3.10586E-15
accept	78	179	257	-64.8017	8.28272E-16
agree	1	80	81	-118.7780	1.17117E-27
realize	39	210	249	-168.4420	1.61960E-38
see	319	633	952	-188.9880	5.28641E-43

(12). This is a stark contrast with Japanese essays which only include 'In conclusion' in their final paragraphs. Several fixed expressions with the verb *mention* are also obvious in various patterns such as 'As I mentioned above,' 'According to the case I mentioned above,' 'As mentioned,' and 'From the standpoint mentioned.' These are useful transitional expressions to connect sentences or paragraphs. This leads to the assumption that Chinese learners of English are perhaps frequently taught these fixed forms and the way to use them appropriately.

### Distinctive Verbs Used by NGS Compared with NES

Some unique words revealed by QCT, which stand out as being particular to German learners, include *tell*, *imagine*, *think*, *complain*, *notice* and other basic verbs. Their statistical significance is testified by LL as listed in Table 5 ( $p < .0001$ ).

None of the noticeable academic verbs are identified in comparison with NES, and this study pays a particular attention to the colloquial expressions obtained from the concordance lines for *tell* and *imagine*.

In terms of *tell*, the examples include 'I must tell you about my opinion' or 'Let me tell you why.' These sentences reveal that students may try to use first person pronouns in conjunction with *tell* to interest the readers in their statement in the beginning of the



paragraph. This also highlights the speech-like nature of learner writing. The verb *imagine* include such expressions as 'I can't imagine,' '(Just) imagine that,' 'I could imagine,' 'Can you imagine ... ?' which are prevalent. In the same way as *tell*, German learners of English are writing as if they are talking even in argumentative essays. This similar feature is also found among other European nationalities (Granger, 1998b).

Table 5

*Frequencies of Verbs Characteristically Used by NGS against NES and their LLRatioStatistics*

Reporting Verb	NGS	NES	Total	Signed LL	Probability
tell	162	148	310	66.39638	3.68779E-16
forget	63	31	94	54.03930	1.96519E-13
remember	71	42	113	51.21948	8.25951E-13
imagine	46	16	62	50.83294	1.00571E-12
think	306	430	736	44.69848	2.29836E-11
know	254	355	609	37.36334	9.80479E-10
find	218	315	533	28.10362	1.14991E-07
complain	19	5	24	24.44574	7.64323E-07
notice	31	17	48	23.97362	9.76649E-07
realize	40	210	250	-27.9314	1.25692E-07
determine	2	64	66	-33.2379	8.15473E-09
allow	52	265	317	-33.4803	7.19884E-09
argue	76	368	444	-42.3288	7.71464E-11
believe	0	62	62	-44.9568	2.01431E-11
state	0	128	128	-93.1413	4.86907E-22

**Discussions for the Findings obtained from NCS and NGS**

In this study, the choice of the Chinese corpus derived from the assumption that their texts may reveal similar features to the Japanese texts; however, the results of the analysis did not confirm this assumption. The results of QCT and LL statistics show that the Chinese essays contain a variety of academic verbs in a native-like manner. The words markedly noticeable in the Chinese texts are *indicate*, *argue*, and *claim*. These verbs are used with third person subjects, which imply strong support for one's opinion given by a third party. Suasive verbs such as *suggest*, *recommend*, *urge*, and *propose*, on the other hand, are often used to propose personal opinions in the essay. Another feature is a variety of fixed expressions identified especially for *conclude* and *mention*. This might be attributed to the education Chinese students received; however, no evidence is given on this matter. One of the reasons why NCS may deal with diverse reporting verbs in a native-like manner might be related to the fact that the data collected in the corpus is not only from Chinese English learners in mainland China (EFL) but also from Chinese students in Hong Kong (ESL). The data from the latter subjects may highly influence the results of the NCS data; however, this assumption needs more investigation.

German learners in contrast display features of speech for their writing. NGS use *tell* and *imagine* to catch the attention of the readers, giving their essays more of a conversational style. This result underscores the finding by Lorentz (1999). It is interesting to note that similar features are identified in other European nationalities (Granger & Rayson, 1998; Gilquin & Paquot, 2008). Granger (1998b) found a speech-like characteristic in French writers, ascribing it to pedagogical factors which focused on speaking. Gilquin and Paquot (2008) refer to L2 writers' confusing signals of register and attributed it to four possible reasons; the influence of the spoken medium, the influence of the mother tongue, the influence of teaching and the effect of developmental factors. German essays contain *think* most conspicuously although the frequency is far less than the Japanese essays. This indicates what Granger (1998b) calls "cover-all" *think* identified among European nationalities; however, there is still room for further investigation into related research.

### Pedagogical Implications

This section, based on the result of this study, will discuss the pedagogical issues. The results showed that learners have a limited lexical knowledge and an underlying confusion about registers. In addition, there is a possible lack in the knowledge of rhetorical organization of academic prose. The important issue to note here is that these results are the most apparent in the Japanese texts among the four language nationality groups from a corpus of relatively advanced learners. Hinkel argued in her study (2003) that learners' unsophisticated text usage was obvious even among advanced learners who would have been exposed to substantial amounts of reading and experience with writing in academic contexts during the course of study at university. This means that mere exposure to the target text is not enough. Much research has indicated that a substantial and advanced L2 proficiency in lexis and grammar may not be possible to achieve without explicit, focused, and consistent instruction (Celce-Murcia, 1991, 1993; Celce-Murcia & Hilles, 1988; Coady & Huckin, 1997; N. Ellis, 1994; R. Ellis, 1984, 1990, 1994, 1997, 2002; Hinkel, 1992, 1997, 2002; Nation, 1990, 2001; Norris & Ortega, 2000; Schmidt, 1990, 1995, to mention just a few, cited in Hinkel, 2004). Some possible instruction suggestions are presented below.

One is corpus-driven instruction. The knowledge gained from corpus-analysis can be a useful resource for teaching. Thompson and Ye (1991) extracted reporting verbs from published journals in diverse fields and classified them for use. Bloch (2010) created a small corpus of 540 sentences from *Science* journal, which include 27 reporting verbs for examination, making them into a database of sentences used to create teaching materials for academic writing. The authenticity of this data cannot be underestimated because the sentences included in the corpora were gathered from materials that native speakers of English actually wrote. Teachers are able to use this corpus-generated database in the classroom, for example, presenting students with alternative words for use. More specifically, alternative words to the verb *think* such as *consider*, *debate*, *deliberate*, *explore* may be found in the database. In order to show the semantic usage, teachers could refer to concordance lines and discuss the meaning or usage in the classroom. Granger (1998a) mentioned that data-driven learning by concordance encourages 'conscious-raising' (Ellis, 1991; James, 1992, cited in Granger, 1998a), meaning that learners are able to be aware of the features that require

attention. By actually using those words in their writing, learners will be able to internalize the meaning and usage of the words in context.

Another approach is genre-based writing instruction. Genre theory has had a major impact in the field of EAP (e.g. Swales, 1984; Dudley-Evans, 1986; Hopkins & Dudley-Evans, 1988, cited in Hyland, 1999). In genre approach, teachers seek to offer students an explicit understanding of how texts in target genres are structured and why they are written in the way they are (Hyland, 2003). Each genre of texts such as argumentative, expository, narrative or informative has a specific audience, purpose, text type, structure and lexis, and this information should be presented to the students at the beginning of a study course in writing. Each paragraph functions in terms of ‘moves’ and ‘stages’ and this feature can be discussed with the students so that they will be able to identify how to construct a text appropriate to the genre. This approach has been used by researchers. Hyland (1999) used genre approach in the teaching of argumentative essays to high school students in Papua New Guinea, showing effective results in their writings. Bithener and Turner (2008) reported the effectiveness of teaching approach in the writing of literature reviews. They described a 15-hour unit of teaching as part of a six-day intensive graduate writing course. The study found clear evidence of improvement based on genre-induced writing pedagogy. This method seems to be effective especially for Japanese learners, whose rhetorical structure and register differ from native speakers of English, to acquire text construction skills and build academic vocabulary.

## Conclusion

The aim of this study was to reveal the characteristics seen in the writings of learners of English through a comparative analysis, focusing on reporting verbs. The investigation was conducted to observe whether the appropriate range of verbs to express opinions is used and how their usages differ between L1 and L2. The research was conducted both quantitatively and qualitatively. The results revealed remarkable characteristics among the four nationalities. Japanese texts, in particular, showed a much simpler structure with limited lexical variety and a speech-like nature, and they overwhelmingly employed the word *think*. Based on the findings, the necessity of explicit, focused, and consistent instruction for academic text and vocabulary use is obvious for Japanese learners, and the two suggested pedagogical approaches, which are the corpus-driven conscious-raised approach and genre-based process writing, were discussed.

Upon reflection, there are some considerations for further study. Firstly, the selection of reporting verbs needs to be re-examined. Classification of these verbs for this type of research changes according to the type of writing to be investigated. Academic papers, for instance, contain more specific academic vocabulary, which should be examined further from the perspective of ESP. Secondly, as the Chinese data contained essays written by students in Hong Kong, which may have affected the sample, only the data from mainland China should be used to see what results may occur. Thirdly, the German data needs to be reexamined from the viewpoint of the rhetorical features of writing to determine whether or not their essays are written with an informal tone. Fourthly, pedagogical approaches that effectively help Japanese learners of English in their academic writing need to be investigated from other ESL or EFL countries so that Japanese learners will be able to write relevantly according to the genre.

This study represents only a fraction of learner corpora research. However, the present writer hopes that the findings here deliver important insights for future research in related fields as well as English education in Japan, especially for academic writing.

## References

- Aijmer, K. (2002). Modality in advanced Swedish learners' written interlanguage. In S. Granger, J. Hung & S. Petch-Tyson (Eds.), *Computer Learner Corpora, Second Language Acquisition and Foreign Language Teaching* [Language Learning and Language Teaching 6] (pp. 55–76). Amsterdam & Philadelphia: John Benjamins Publishing Company.
- Bichener, J., & Turner, E. (2008). *An Approach to Teaching the Writing of Literature Reviews*. Retrieved from [www.zeitschreiben.eu](http://www.zeitschreiben.eu).
- Bloch, J. (2010). A concordance-based study of the use of reporting verbs as rhetorical devices in academic papers. *Journal of Writing Research*, 2(2), 219–244.
- Carlson, S. (1988). Cultural differences in writing and reasoning skills. In A. Purves (Ed.), *Writing across languages and cultures: Issues in contrastive rhetoric* (pp. 227–260). Newbury Park, CA: Sage.
- Charles, M. (2006). Phraseological patterns in reporting clauses used in citation: A corpus-based study of theses in two disciplines. *English for Specific Purposes*, 25, 310–331.
- Goto, K. (2006). Go-Tagger (ver 0.7) [Computer Software]. Tokyo, Japan: Setsunan University. Retrieved from <http://web4u.setsunan.ac.jp/Website/GoTagger.htm>
- Granger, S., & Rayson, P. (1998). Automatic profiling of learner texts. In S. Granger (Ed.), *Learner English on Computer* (pp. 119–131). London and New York: Addison Wesley Longman.
- Granger, S. (1998a). Learner corpus data in the foreign language classroom: form-focused instruction and data-driven learning. In S. Granger (Ed.), *Learner English on Computer* (pp. 199–209). London and New York: Addison Wesley Longman.
- Granger, S. (1998b). The computer learner corpus: a versatile new source of data for SLA research. In S. Granger (Ed.), *Learner English on Computer* (pp. 3–18). London and New York: Addison Wesley Longman.
- Gilquin G., & Paquot M. (2008). Too chatty: Learner academic writing and register variation. *English Text Construction*, 1(1), 41–61.
- Hasselgren, A. (1994). Lexical teddy bears and advanced learners: a study into the ways Norwegian students cope with English vocabulary. *International Journal of Applied Linguistics*, 4(2), 237–258.
- Hinkel, E. (2002). *Second Language Writer's Text: linguistic and rhetorical features*. New Jersey: Lawrence Erlbaum Associates, Inc.
- Hinkel, E. (2003). Simplicity without elegance: features of sentences in L1 and L2 academic texts. *TESOL Quarterly*, 37, 275–301.
- Hinkel, E. (2004). *Teaching Academic ESL Writing: Practical Techniques in Vocabulary and Grammar*. New Jersey: Lawrence Erlbaum Associates, Inc.
- Hyland, K. (1999). Academic Attribution: Citation and Construction of Disciplinary Knowledge. *Applied Linguistics*, 20(3), 341–467.
- Hyland, K. (2003). Genre pedagogy: A social response to process. *Journal of Second Language*

*Writing*, 12(1), 17–29.

Hyland, K. (2005). *Metadiscourse*. London: Continuum.

Ishikawa, S. (2008). *Eigo Corpus to Gengo Kyoiku: Deta to Shite no Tekusuto [English Corpus and Language Education: Texts as data]*. Tokyo: Taisyukan.

Ishikawa, S. (2010). Nihonjin Eigogakushusha ni yoru Chukangengo no Goiunyou: Gakushusha Corpus CEEAUS o mochiita Tasoutekitaishou Chukangengo Bunseki [A Study of Interlanguage of Japanese Learners of English.: Analysis of Interlanguage of Several Language Nationalities Using the Learner Corpora CEEAUS]. In H. Kishimoto (Ed.), *Gengo no Taishou [Objects of Language Research]*. (pp. 217–231). Tokyo: Kuroshio Publishing Co.

Kobayashi, Y. (2009) NS/NNS Tekisutobunruimoderu ni motozuku Nihonjin Eisakubun no Tokucho Chushutsu [Features of English Essays Written by Japanese Based on Text Classification Model]. *Jinbunkagaku to Konpyuta Shinpojiumu Ronbunshu – Digital Humanities no Kanousei [Research Presented in Symposium on Humanities and Computers-Possibility of Digital Humanities]*. (pp. 261–268). Tokyo: Information Processing Society of Japan.

Laurence, A. (2000). AntConc (ver 3.2.1.) [Computer Software]. Tokyo, Japan: Waseda University. Retrieved from [http://www.antlab.sci.waseda.ac.jp/antconc\\_index.html](http://www.antlab.sci.waseda.ac.jp/antconc_index.html).

Leki, I. (1991). Twenty-five years of contrastive rhetoric: Text analysis and writing pedagogies. *TESOL Quarterly*, 25(1), 123–143.

Lorenz, G. (1999). *Adjective Intensification. Learners versus Native Speakers. A Corpus Study of Argumentative Writing*. New York: Rodopi.

McCrostie, J. (2008). Writer Visibility in EFL Learner Academic Writing: A Corpus-based study. *ICAME Journal*, 32, 97–114

Moriwaki, K. (2011). A Comparative Study of CEO Letters between U.S. and Japanese Companies. In *On Language and Language Education Vol 4* (pp. 63–92). Kyoto: Graduate School of Language Education and Information Science. Ritsumeikan University.

Nakamura, J. (2006). Freq Analysis Toolkit (Version 4.0) [Computer Software]. Kyoto, Japan: Ritsumeikan University.

Neff, J., F. Ballesteros., E. Dafouz., F. Martínez., & J-P. Rica. (2003). A Contrastive Functional Analysis of Errors in Spanish EFL University Writers Argumentative Texts: a Corpus-based Study. In E. Fitzpatrick (Ed.), *Corpus Linguistics Beyond the Word: Corpus Research from Phrase to Discourse* (pp. 203–226). New York: Rodopi.

Petch-Tyson, S. (1998). Writer/reader visibility in EFL written discourse. In S. Granger (Ed.), *Learner English on computer* (pp. 107–118). London and New York: Longman.

Quirk, R., Greenbaum, S., Leech, G. & Svartvik, J. (1985). *A Comprehensive Grammar of the English Language*. London: Longman

Read, J. (2000). *Assessing Vocabulary*. Cambridge: Cambridge University Press.

Ringbom, H. (1998). Vocabulary frequencies in advanced learner English: a cross-linguistic approach. In S. Granger (Ed.), *Learner English on Computer* (pp. 41–52). London and New York: Addison Wesley Longman.

Takami, S. (2004). A corpus-driven Identification of Distinctive words: ‘tablod Adjectives’ and ‘broadsheet Adjectives’ in the Bank of English. In S. Saito, J. Nakamura & S. Tamazaki

- (Eds.), *English Corpus Linguistic in Japan* (pp. 115–135). Amsterdam: Rodopi.
- Tsukamoto, S. (2006). KWIC Concordance for Windows (ver 4.7) [Computer Software]. Tokyo, Japan: Nihon University. Retrieved from [http://www.chs.nihon-u.ac.jp/eng\\_dpt/tukamoto/kwic4.html](http://www.chs.nihon-u.ac.jp/eng_dpt/tukamoto/kwic4.html)
- Thompson, G. & Ye, Y. (1991). Evaluation in the Reporting Verbs Used in Academic Papers. *Applied Linguistics*, 12(4), 365–382.

### Acknowledgements

I would like to express my gratitude to all those who gave me the possibility to complete this thesis. I am deeply indebted to Prof. Junsaku Nakamura, my supervisor and the former professor from Ritsumeikan University whose continuous help and stimulating suggestions helped me a lot. I would also like to thank Dr. Hideyuki Taura for his encouragement, Mr. Daisuke Abe for correcting my thesis, and anonymous referees whose comments enabled me to complete this work.

### Notes

1. Eight verbs, i.e., *confide*, *conjecture*, *enjoin*, *entreat*, *foretell*, *ordain*, *retort* and *sense*, which no nationality uses, are deleted since they are not meaningful; thus, the total number of reporting verbs for analysis is 142 words in this study.
2. The verbs which appear only in one particular corpus are combined together as a group identified by nationality in Table 2 since they are given exactly the same quantities by QCT, eventually making them overlap one another when plotting the results. For example, “allege” (occurring 3 times), “ascertain” (1), “beg” (2), “calculate” (7), “commend” (2), and “decree” (1), which appear only in NES texts, are grouped together as “NES” in the sample distribution. In the same way, there are 3 verbs unique to Chinese, i.e., “boast” (1), “deduce” (2), and “stipulate” (3), and they are combined together as NCS. NGS contains the verbs like “fancy” (2), “forecast” (1), “presuppose” (2), “prophesy” (1), and “vow” (1). There are no verbs uniquely used by NJS.
3. A Log-Likelihood Ratio Statistic is a measure used for determining whether or not the words appearing in two corpora are used in a significantly different way, eventually making it possible to find out the keywords of a particular corpus in relation to a reference corpus as demonstrated by WordSmith tools. Collocations in a particular corpus or a text can be picked out by this method, too. (See Takami, S., 2004, p. 120 for detail.)