

# A new parallel corpus approach to Japanese learners' English, using their corrected essays

Nozomi Miki

u563877d@ecs.cmc.osaka-u.ac.jp

Graduate School of Language and Culture, Osaka University, Japan

## Abstract

This research introduces unique parallel corpora to uncover linguistic behaviors in L2 argumentative writing in the exact correspondence to their appropriate forms provided by English native speakers (NSs). The current paper targets at the mysterious behavior of I think in argumentative prose. I think is regarded as arguably problematic and controversial in L2 writing community; the overuse of this marker underscores the points, confusing logical development, while the rubric of argumentative writing tests such as TOELF and IELTS requires test-takers to write their personal views. Actually, I think is favoured as a booster in academic writing of particular disciplines. Our approach resolves this, investigating how I think was used in Japanese students' writing and corrected by NSs, specifically, how I think was deleted or kept by NSs. This is a promising method and immediately applicable to classroom teaching, where university students can receive feedback from our unique parallel corpora.

## Introduction

A learner corpus—a corpus from texts written by non-native learners of English—was a new departure, compared to corpora in general, but gives us different insights into learner English: general corpora (in a corpus term, comparative corpora) describe what real English is like, while learner corpora tell us what learner English is like. General corpora help to remove deviant, odd, artificial English from ELT materials and raise learners' awareness of the differences between authentic English and English which they are exposed to in class or in textbooks. On the other hand, learner corpora inform us of what learners find difficult about English writing and what is non-native-like in their English, including their fossilized errors (Granger, 1998, pp. 3–18; Granger, 2002, pp. 3–33; Hunston, 2002, pp.206–216). To put it another way,

learner corpora can answer a frequent question from learners: “Why is my English foreign-sounding, though it is grammatically correct?”

Studies of learner corpora have explicitly addressed the characteristics of learner English and the possible factors; it is regarded as “dull, repetitive, unimaginative” and “verbose” (Ringbon, 1998, p. 50). Gillard and Gadsby (1998, p. 161) investigated the frequencies of synonyms such as *big*, *enormous*, *massive* and *huge*, concluding that learners are highly likely to use the same words in expressing particular ideas. Similarly, de Cook et al. (1988, pp. 71–73), based on the recurrent words and combinations in the learner spoken corpus, stressed that learners did use formulaic expressions but the recurrent ones.

The way of characterizing learner English depends on the overuses/underuses of particular words by learners in comparison with control corpora or reference corpora (i.e., the data sets of NS essays). The overuses/underuses are such relative concepts that results are subject to change, according to control corpora, or reference corpora; the problem is that a learner corpus has different overuses in comparison with different reference corpora (Baker, 2004; Miki, 2009). There are no correspondence between learner corpora and control corpora. NNSs and NSs were supposed to write freely about specific topics, so that word numbers per essay varies, as in ICLE (the International Corpus of Learner English) and LOCNESS (the Louvain Corpus of Native English Essays). LOCNESS, which consists of American and British student essays, is probably best-known as a reference corpus for learner corpora and has been a driving force of the considerable number of L2 writing studies. LOCNESS is controlled in terms of register (i.e., argumentative writing), writers’ first languages (American/British English), and age (university students). However, the reality is that LOCNESS is a miscellany of argumentative writing by NS university students. Most of the texts were written as class assignments, using reference tools such as library books, newspapers, and dictionaries, while some are included timed essays such as term-end examinations. LOCNESS also covers seemingly academic topics about French literature, culture-specific topics such as “fox hunting” and “a single Europe: a loss of sovereignty for Britain,” “gun control,” and “prayer in schools,” which may not be issues in other countries. They make a sharp contrast with the general topics of learner corpora which does not require any specialized knowledge. The information in LOCNESS is rich, since the writers were allowed to consult reference books, leading to uses of academic or foreign words, while it includes careless grammatical and orthographical errors (e.g., using “&” despite formal writing). As a result, the length of essays varies even within LOCNESS itself. LOCNESS casts doubts over comparability with learner essays, even though the statistics

solve the problem with different sizes. Due to this, the recent reference corpora for learner corpora are strictly controlled with variables including time and topics rather than age (cf. Sugiura, 2008). Seriously, without the correspondence, there is no instant answer to a frequent question from learners and teachers, "How would native speakers put these learners' expressions in English?" Thus, the overuse/underuse approach informs the L2 writing community of learners' foreign-sounding, non-native-like characteristics but the fact is that it is not enough for pedagogical application.

The current paper presents a new approach to learner English, thus, proposes to view it from another angle by applying an idea of parallel corpora, which is exploited in translation studies; comparing a target language with a source language on a line basis. Instead, our parallel corpora were specially made by aligning NNS English with the corresponding NS English line by line so as to pin out exactly what NS English corresponds to the NNS uses. Our approach to L2 writing gives NNS writing teachers possible alternatives to learners' errors and ways to correct them with a line-by-line correspondence between the parallel texts in classrooms. This method also lets L2 writing practitioners find out NNS unhidden uses from their writing corrected by NS teachers (i.e., phraseology) from the comparison between the NNS writing and their corrections by NSs.

In order to demonstrate this approach, *I think* was chosen as a target for this study; there are several previous studies which indicated that this phrase is frequent in argumentative writing not only of Japanese learners of English but also of other backgrounds (Petch-Tyson, 1998, pp. 114–115; Ringbom, 1998, pp. 43–44) and problematic to learners of English. McCrosite (2008) investigated visibility (i.g., uses of pronouns) in his learner corpus of about 200,000 words from 333 argumentative essays by Japanese university students and indicated the overuse of *I*, specifically, *I think*; his Japanese learners had a much higher frequency of this than the European students of Petch-Tyson (1998). Oi (2002, pp. 64–67) mentions that the overuse of *I think* can be interpreted as "lack of confidence"; it is presupposed that writers express their opinions in argumentative writing and then do not bother to say so with *I think*. Similarly, Ando (2006, p. 158) indicates that *I think* is an obstacle to interpreting argumentative essays as in the TOEFL writing section. Kobayashi (2009), which compared Japanese ICLE and LOCNESS, argued that this Japanese learners' cliché was due to the lack of lexical variations of boosting phrases, suggesting L1 transfer; there are so many English translations of a single Japanese phrase, *omou* (*I think*) in a Japanese-English dictionary.

What is lacking in the previous studies of *I think* is that they do not tell us how learners employ this phrase, where it is not appropriate. Oi (2002) illustrated a NS essay together with a NS but failed to make a close examination of the linguistic behavior of this phrase from a number of NNS/NS texts. Kobayashi (2009) just indicated the high visibility of Japanese learners in comparison with NS students, based on the quantitative, statistic analysis. In addition to the overuse of this phrase, McCrostie (2008, p.109) stated two characteristics of his students' *I think*: first, his Japanese students employed this phrase in order to express an opinion in contrast to NSs who mentioned the personal past experiences; secondly, Japanese learners' frequently occurring position of *I think* was at the beginning of sentences, while European students tended to place it at the end of sentences. This seem not to be enough to inform EFL teachers and learners of the particular behavior of *I think* in argumentative prose, because NSs do use *I think* in the clause-initial positions in argumentative essays as a booster, though it is not so frequent as learners.

According to Miki (2008, pp. 54–55), *I think* was more frequent in the collection of TOEFL model essays from 17 published textbooks (97,305 words), which were written or proofread by NSs and highly idealized texts, than LOCNESS by non-professional American student writers, which comprises errors, probably because it was elicited by the rubric of TOEFL (e.g., *Why do you think people attend college or university?*). The same is true for academic writing, whose register is argumentative. According to Hyland (2000), *I think* function as a booster which emphasizes a writer's position wither certainty especially in academic writing of applied linguistics and philosophy, as in:

- (1) a. ***I think*** there is a correct answer to the question 'Does God exist?' (Philosophy) (Hyland, 2000: 123)
- b. ***I think*** it works something like this: suppose we start with a new, just-assembled ship S ... (RA: Phil) (Hyland, 2002, p. 1103)

Hyland (2000, p. 123) mentioned that the use of the first person pronouns marks "an overt acceptance of personal responsibility for a judgment" and represents "a confident and expert mind in full control of the material." McCrostie (2008, pp. 110–111) argued that L2 writers and professional academic writers use *I think* in different manners to serve different purposes, casting a doubt whether novice or intermediate level EFL learners represent "a confident and expert mind in full control of the material" just because they use the same expressions.

However, it is important to note that the use of *I think* differs between soft disciplines and hard science; since philosophy and applied linguistics place an emphasis on writers' knowledge, logics, and teaching experiences, personal comments, illustrated

by *I think*, are acceptable. On the other hand, hard science places a vital importance on data and experiments which contributes to objective evidence, resulting in almost no room for *I think*. Given social but general topics of L2 writing, it is expected that L2 writers use *I think* to express a personal opinion similarly to the uses of *I think* in soft disciplines. Whether learners are not fully confident of the materials would be another matter.

Rather, it is necessary to consider that *I think* is different in the degree of certainty from other boosters; Hyland (2000) mixed up *I think* and boosters of a high degree of certainty (e.g., *I am convinced*). Biber et al. (1999) suggest that *I think* bears the different degree of certainty from others.

... the verb *think* controlling *that*-complement clause not only marks the degree of certainty (being less certain than verbs like *know* but more certain than verbs like *suspect*), but also indicates the source of knowledge.

*Since last year I think they have improved.* (news) (Biber, et al., 1999, p. 972)

From the above, it seems that *I think* work as a booster to mark authority and a hedge to send a writer's interpretation modestly. This fact has been ignored in the previous studies of *I think*.

This research aims to locate exactly what Japanese learners' *I think* corresponds to in the corrections by NSs, using the parallel corpora. Some of *I think* by Japanese learners were deleted by NSs, others survived in their corrected texts. This paper will then clarify their actual uses of this controversial phrase in argumentative writing, arguing that *I think* in argumentative writing is not a matter of quantity. Lastly, I will suggest how L2 writing teachers instruct their students about their familiar phrase, *I think*.

## Methodology

For this research, I made parallel corpora from NICE (Nagoya Interlanguage Corpus of English, 207 texts, 69,875 words), a dataset of argumentative essays by the Japanese and undergraduate or graduate students (Sugiura, 2008). In order to do SLA research NICE is highly controlled in terms of tasks, proficiency levels and topics.

In NICE 207 participants wrote an essay within an hour without using dictionaries (i.e., it was not homework) so as to see how many words learners can produce within the fixed time, while in ICLE, there were some homework essays in which students may have used dictionaries. In learner corpora, the definition of proficiency levels are not so clear; ICLE regards "advanced" students as third-year or fourth-year students with major in English literature or linguistics, while Lorenz (1999, p. 10) defined it as

“learners who are generally expected to have mastered the basic rules and regularities of the language they are learning.” In contrast, NICE included the score of TOEIC and/or TOEFL of, if not all, some learners (74 out of 207 participants), so as to discover the developmental stages of their writing. ICLE comprises 922 topics, which makes it difficult to sort out the essays under the same or similar themes, while NICE selected 11 topics which are not culture-specific but universal ones (see Figure 1).

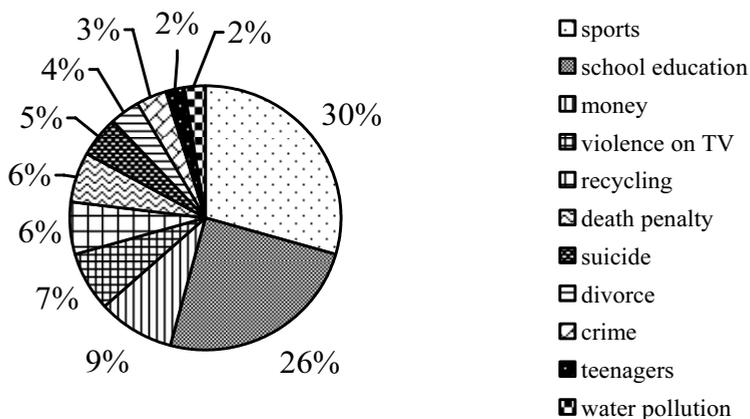


Figure 1. Themes in NICE

As Figure 1 shows, the themes are single phrases (e.g., “sports”) without specific rubrics and general, everyday’s matters, far from scholarly, academic ones. The learners received the short lecture of structures of argumentation such as “introduction,” “body,” and “conclusion” beforehand so that the lack of knowledge about argumentative prose would not influence the result. They were asked to begin essays with their own title about a theme and developed it within an hour with no reference tool.

Besides the controlled variables and their own NS corpus, what makes NICE distinct from other learner corpora is corrections attached to each essay on the line basis. The students’ essays were corrected by a proofreading company in Japan to keep the high quality of correction by NSs; NS proofreaders were well-trained and qualified professionals by the institution. They were asked to use learners’ words in correction as long as they are appropriate, though they preferred formal phrases which fit into

argumentative writing (e.g., *therefore* instead of *so*). The exact number of proofreaders engaged was not released by the proofreading company and NICE but from the NS comment of corrected essays, a proofreader was in charge of several texts.

The learners' texts (JPN) were saved on a sentence basis, followed by NSs' corrected ones (NTV) and optional comments (COM), as seen in Figure 2. From this dataset, I made two independent corpora including Japanese learners' essays and those corrected by NSs respectively, aligning each line for correspondence. In case that a sentence in JPN was corrected into two sentences by NSs or vice versa, these two lines were combined into a single line for exact correspondence.

*JPN030:	Many people are coming to this city from many parts of the world for the job, business or their own purposes.
%NTV:	Many people are coming to Nagoya from many parts of the world for business or personal reasons.
%COM:	Job and business do not really seem to be different things.

**Figure 2. The original text in the Japanese learner corpus of the NICE**

There were 74 texts out of 207 with the score of English proficiency tests (TOEIC and TOEFL). In order to see different behaviors, according to proficiency, I selected 60 essays of learners who took the tests within two years before they produced them for NICE so as to avoid the change of their proficiency over the time. The sixty JPN texts and the corresponding NTV were divided into three on the basis of TOEIC score (20 texts each). Just in the same way, the corresponding essays in NTV were named "high NTV," "middle NTV" and "low NTV" for reference (see Table 1 and See Table 2).

**Table 1. The division of proficiency levels, based on the TOEIC score**

Levels	TOEIC Score	Actual score
Low	less than 550	380-550
Middle	600-800	620-760
High	more than 860	880-940

Table 2. The wordlists of JPN and NTV at different levels

	JPN (JPN essays)			NTV (corrections by NSs)		
	Low	Middle	High	Low	Middle	High
Tokens	4,987	7,451	8,941	5,239	7,258	8,462
Types	941	1,295	1,634	1,036	1,356	1,602
STTR	49.77	53.92	56.57	54.59	57.15	56.63
Length of sentence	10.56	13.44	13.08	12.32	12.96	16.92

Note: 20 texts at each level

Admittedly, our graded corpora were not large but relatively small. It should be noted, however, that learner corpora are not so large as comprehensive mega corpora such as BNC and the Bank of English; for example, Lorenz (1999: 17) compiled small corpora such as German teenagers' corpus containing 130,730 words and German graduates' corpus comprising 72,031. According to O'keeffe et al. (2007, p. 4), small corpora would suffice if researchers investigate quite common words and structures. *I think* is such a frequent stance marker (Biber, et al., 1999, p. 373) and not a culture- or topic-specific expression; the current size would be sufficient for the purpose of this investigation.

There are merits in such small corpora. A corpus pioneer, Sinclair who advocated large size corpora, acknowledged four advantages of small corpora in his latter research life (Sinclair, 2001): (1) researchers can directly interpret evidence; (2) it is easy to discover particular features in small specialized texts; (3) differences can be found by comparison irrespective of corpus sizes; (4) small corpora offer tail-ordered resources for classroom use easily. Our investigation took advantage of them.

Since the sizes of JPN and NTV vary, normalization (henceforth, NF) is used but only for illustration; it distorts low frequencies. To see whether difference in a frequency of *I think* between JPN and NTV are due to chances or not, the log-likelihood ratios (LL) was employed. Dunning (1993) indicates that this statistics excels another statistical index, a chi-square test in dealing with different sizes of data. The significance levels were determined, according to the website of the log-likelihood ratio by University Centre for Computer Corpus Research on Language ( $p < .05$  for a critical value of 3.84;  $p < .01$  for 6.63;  $p < .001$  for 10.83).

For this research, a software tool, *ParaConc* (Barlow, 2004) was used to search a learner corpus for a particular word or phrase on a parallel with a corpus which consists of their corrections. Using this software, the corresponding target items

were obtained in the two directions. When the JPN corpus (the source data) was searched for a particular word, the corresponding items were obtained from the corrected essays (the target data) including the same ones, the alternatives or deletions. Conversely, when the corrected essays (the source data, in turn) were searched for the same word, the corresponding patterns were displayed from the learners' original essays. The next section will reveal not only **how** the Japanese learners used *I think* in persuasive prose but also **how** this phrase should be used from the corrected ones by NSs, and uncover the unknown, detailed usages of *I think* by NSs in argumentative writing.

## Results and discussion

### *The frequencies of I think in JPN and NTV*

As the previous studies indicates, *I think* was significantly more frequent in JPN than in NTV (see Table 3). There were 84 instances of *I think*; 47 were deleted and 37 were kept in the NS correction. The NS proofreaders deleted *I think* by JPN from 30 instances to 16 at the low level, from 28 to 14 at the middle level and from 26 to 7 at the high level, all of which mark significance ( $p < .05$ ). The NSs were asked to keep as many original words by learners as possible in correcting their essays but actually erased them drastically.

However, it should also be noted that not all of the instances of *I think* were deleted. Most of the deletion was due to the hedging uses. I will take up noticeable examples which are not in considerable number but are pedagogically suggestive. Specifically, I will show in what environment and how they were removed and kept in the NS correction.

**Table 3. The frequencies of I and I think at different levels**

Levels	LL of <i>I think</i>	Freq of <i>I think</i>	Freq of <i>I</i>	Levels	LL of <i>I think</i>	Freq of <i>I think</i>	Freq of <i>I</i>
low JPN	156.56	30 (3.36)	204 (22.82)	low NTV	58.52	16 (1.89)	200 (23.64)
middle JPN	131.24	28 (3.76)	139 (18.66)	middle NTV	57.56	14 (1.93)	118 (16.26)
high JPN	118.44	26 (5.21)	126 (25.27)	high NTV	27.57	7 (1.34)	78 (14.89)

Notes: LL stands for the log-likelihood ratios. The number of ( ) is normalized frequencies per 1000. *I think of*, *I think about*, and the quoted one were excluded from frequencies.

### ***Deleting or replacing I think with another phrase***

*I think* was deleted especially due to the logical contradiction. In Table 4, the Japanese learners slotted *I think* into rigid, logical relations containing *if* to make their evaluations.

**Table 4. The deletion of I think in the conditional statement**

	The NNS writing	The NS correction
1	<i>If</i> there are such a people remaining, <i>I think</i> we can't keep the recycling system well even many people tries. [high JPN]	<i>Despite</i> these changes in the garbage collection system, it is not working well.
2	I discussed about "learning English from early age" from three points, and <i>if</i> I say in one word, <i>I think</i> we shouldn't start learning English from an early age. [high JPN]	Having examined the problems that can be caused by learning English from an early age and thought through a learning technique we might apply, <i>I believe</i> that we should not start learning English at an early age.
3	<i>If</i> such place will expand more and more, <i>I think</i> it will give good effect to school education. [low JPN]	<i>If</i> this happens more and more frequently, it will have a positive effect on school education.
4	So, <i>even if</i> English is not our mother tongue, <i>I think</i> it will help us to follow the steps what the English people do to learn their mother tongues. [high JPN]	<i>Even though</i> English is not our mother tongue, <i>I think</i> we would benefit from following the steps by which native English speakers learn their mother tongue.

Note: 4 occurrences (High, 3; Middle, 0; Low, 1)

In the NS correction *I think* and/or *if* were eliminated or replaced with another expressions. There is no space for any personal evaluations to make in the objective, conditional relations such as *if ... and then ...* Considering that *I think* is a manifestation of personal certainty, statements including this phrase are not strictly logical or well-grounded or are not so certain; specifically, "less certain than a verb *know*" (Biber et al, 1999). This marker is a hedge so that it was not compatible with other epistemic markers of conviction, or boosters. See another type of logical errors in Table 5.

A logical consequence based on facts does not give any room for a personal intruder such as *I think*. Otherwise, this personal marker would divert and downgrade the reasoning. In fact, *I think* was eliminated in Table 5. Notably, there were changes besides NSs' removal of the personal element. In 3 of Table 5, the writers changed

from a personal speaker (e.g., *I can say*) to everyone's observers (e.g., *we can see*), adding general views to it, and, in 4, shifted from a personal thinker (*I think*) to a long-time observer (*I have seen*). The fifth shows us the transformation of the personal summary into an objective, concrete, evidential statement as a matter of fact, changing the human subject into an impersonal subject of a verb, *show*. The sixth pair indicates a shift from a personal thinker (*I think*) to a personal believer (*I ... believe*) to enhance the degree of certainty.

**Table 5. The deletion of I think in the logical relation**

The NNS writing	The NS correction
1 <i>From above, I think</i> the school for teachers is one of the resolve the dangerous of primary school education. [low JPN]	Teacher training is a solution to the dangers of primary school education.
2 <i>And I think we can say</i> the person who gives the death penalty decision to them is also murder. [low JPN]	<i>Therefore, we can say that</i> the person who sentences another to death is also a murder.
3 <i>I think I can say from above</i> that school teacher can give big influence on children's future. [low JPN]	<i>From this example, we can see that</i> a school teacher can have great influence on a child.
4 <i>Then I think</i> many Japanese people recently try to play foreign country and succeed. [low JPN]	<i>I have seen that</i> many Japanese people these days try to play sports in foreign countries, and they succeed.
5 <i>I think it means</i> sumo has the essential that attract people all over the world. [middle JPN]	Sumo's gains in popularity internationally <i>show that</i> it has the ability to attract people all over the world.
6 <i>From above</i> I do not support the existing of the death penalty and <i>I think</i> we should decrease death penalty gradually. [low JPN]	<i>From the above argument, I</i> do not support the death penalty and <i>believe</i> we should gradually stop using it.
7 <i>In result, I think</i> children will have big trust to their teacher. [low JPN]	<i>As a result,</i> children trust their teachers <i>implicitly</i> .
8 For instance, I like spending times talking with my friends and relax over cups of tea with my friends, <i>so I think</i> spending money as social expenses is good to get along with my friends and have my life full. [middle JPN]	For instance, I like spending time talking and relaxing over cups of tea with my friends, <i>so</i> I use money to spend time with my friends and make my life full.

Note: 8 occurrences (High, 0; Middle, 2; Low, 6)

The seventh illustrates how a NS mentioned the third party's state of mind (*children trust ...*) in argumentative writing. Instead of using this marker, the NS proofreader employed a manner adverb tactfully, weakening assertiveness with a covert evaluation, thus, *children trust their teachers implicitly*. The eighth pair of the Japanese writing and the NS correction is another interesting illustration. The NS regarded the evaluative statement of *I think* as a personal experience, or a fact (cf. Petch-Tyson, 1998). Thus Table 5 shows us how writers should manipulate personal views and a matter of uncertainty in argumentation.

Table 6 shows *I think* bears a different degree of certainty from other boosters, resulting in the deletion. *I think* contradicts with *of course*, *probably*, and *the possibility ... is very high*; *I think*, or a personal marker downplayed the probability, which the learners possibly wanted to stress. The NS proofreaders removed all of *I think* in this kind of context towards logical consistency. The second NS correction tells us how best to modify the numbers which writers are not sure of in persuasive prose; using *around* instead of *I think*. It should be noted that these occurred in argumentative writing rather than in conversations, where *I think* may well work as a hedging device in interaction with hearers.

**Table 6. The deletion of *I think* with markers of strong certainty**

	The NNS writing	The NS correction
1	Those three of crime is a very foolish act and <i>I think</i> this is <i>of course</i> the responsibility of their self, but their parents, school and the environment around them too. [middle JPN]	These three crimes are stupid; <i>of course</i> , the responsibility for such behavior lies with themselves, as well as with their parents, teachers, and society around them.
2	Actuary, for example, newspaper's recycling rate is very high, <i>probably</i> 80 percent <i>I think</i> . [low JPN]	The rate newspaper recycling is actually quite high, <i>probably around</i> 80 percent.
3	<i>I think</i> the <i>possibility</i> to trust what teacher says is <i>very high</i> . [low JPN]	The <i>possibility</i> that the child will trust what the teacher says is <i>very high</i> .

Note: 13 occurrences (High, 5; Middle, 2; Low, 6)

There were four instances of *I think* followed by *not*, where a proper sequence should be *I do not think ...* (see Table 7). This kind of error is probably due to the first language, or Japanese, in which it is grammatical to say either *I do not think it is right* or *I think it is not right*, while the latter is ungrammatical in English. More

importantly, the NS proofreaders did not correct the errors into *I do not think* but deleted the whole phrase of *I think*, changing the subordinate clauses containing *not* into the main clauses; otherwise, *I do not think* would personalize the facts including negation, weakening the assertive force.

**Table 7. The deletion of *I think* in a sequence of *I think ... not***

	The NNS writing	The NS correction
1	There were changes in garbage collecting system, however <i>I think</i> that is <i>not</i> working well. [high JPN]	Despite these changes in the garbage collection system, it is <i>not</i> working well.
2	<i>I think</i> that Japanese professors do <i>not</i> expect full comprehension from their students, and that leads to the reason why Japanese classes are easy to get credit. [high JPN]	Professors in Japan do <i>not</i> expect complete comprehension from their students, which is why getting credit in classes in Japan is easy.
3	<i>I think</i> the recycling system is <i>not</i> working well even they changed the system of collecting garbage about 5 years ago. [high JPN]	The recycling system is <i>not</i> working well even though the system of collecting garbage was changed about five years ago.
4	<i>I think</i> the cause of like these accident is the teacher does <i>not</i> know the way of communicating with children. [low JPN]	The cause of incidents like these is that the teacher did <i>not</i> know how to communicate with children.

Note: 5 occurrences (High, 4; Middle, 0; Low, 1)

Instead of *I think*, the NSs selected epistemic auxiliaries and other private verbs to express uncertainty in argumentation, as in Table 8. Modality which the learners intended with *I think* was lexicalized into auxiliaries such as *should*, *must*, *could* and a modal verb, *seem*, adding to objectivity and certainty. The sixth example in particular suggests a different way of giving evidence to the argument from the learner. Both the Japanese writer and the NS proofreader used *because* with the first person singular but *I think* is the mere emergence of ideas, while *I know* means the established knowledge in mind to stress the fact, leading to strong argumentation; again, *I think* is "less certain than verbs like *know*" (Biber et al., 1999).

Lastly, Table 9 indicates that *I think* was removed especially when the writers want to lay a stress on facts and situations to support their views, as indicated by *however* and *but*.

Table 8. The change of I think into other epistemic auxiliaries and verbs

The NNS writing	The NS correction
1 In short, <i>I think</i> the weight of the crime is the weight of the penalty. [low JPN]	In short, the weight of the crime <i>should</i> equal the weight of the penalty.
2 To reject such thing, <i>I think it is important that</i> the teacher is the person who can live in society and can love students. [low JPN]	To prevent this, teachers themselves <i>must</i> be able to function well in society and love children.
3 <i>I think</i> that “their parents want them to entrance private school.” <i>is better</i> . [middle JPN]	<i>I should say that</i> the children’s parents want them to enter private school.
4 <i>I think</i> that even the only one person’s behavior <i>will</i> waste the everyone’s time and energy for taking care about recycling. [high JPN]	The behavior of even one person <i>could</i> waste the time and effort of those who care about recycling.
5 Because of that, <i>I think</i> she never hesitates to say how she thinks, at least in front of me. [middle JPN]	Because of that experience, she never <i>seems</i> to hesitate to say what she thinks.
6 Because <i>I think</i> that I can buy my favorite things by that money. [low JPN]	When I receive my pay, however, I become happy because <i>I know</i> that I can use the money to buy my favorite things.

Table 9. The deletion of I think in the description of supporting situations

The NNS writing	The NS correction
1 There were changes in garbage collecting system, <i>however I think</i> that is not working well. [high JPN]	<i>Despite</i> these changes in the garbage collection system, it is not working well.a
2 My opponent may also say that if people live alone, they can live their own lives, <i>but I think</i> this causes the same problem as above. [high JPN]	Some may say that by living alone, students can live their own lives, <i>but</i> this causes the same problem.
3 <i>I think the only</i> people who miss their family are those who are already close to them, and even people who don’t live alone can still feel the warmth of family life. [high JPN]	<i>The only</i> students who miss their families are those who were already close to them, and we do not need to live alone to discover who important our families are.

Note: 17 occurrences (High, 7; Middle, 10; Low, 0)

*I think* was deleted probably because this element was not necessary for objective description of situations or facts in argumentation, which makes a sharp contrast with surviving *I think* in the next section

### Surviving *I think*

In the previous section, we have seen the deletions of *I think* which play down writers' certainty. In contrast, there were some examples where *I think* survived. *I think* preceded the points expressed by *we need*, *we must*, and *the best way*, as in Table 10. This suggests that the uses of *I think* are not a hedge; the marker does not underscore the argument but mention their own view with some certainty.

**Table 10. *I think* as a booster in argumentative writing**

	The NNS writing	The NS correction
1	<i>I think we really need</i> to think Japanese education over now. [middle JPN]	<i>I think we need</i> to rethink Japanese education.
2	<i>I think</i> the Japanese education <i>necessary</i> to tell students to how to study and enjoy it. [low JPN]	<i>I think</i> that the Japanese education system <i>needs</i> to tell students how to study and how to enjoy it.
3	<i>I think we must</i> justice the person who got the penalty by the same penalty. [middle JPN]	<i>I think we must</i> carry out the execution in a standard way.
4	But <i>I think</i> this is <i>the kind of best</i> way to keep the balance of the internationalization of sumo. [middle JPN]	However, <i>I think</i> this is <i>the best</i> way to maintain the balance of nationalities in the internationalization of sumo.
5	<i>I think</i> that sports is <i>no</i> border. [low JPN]	<i>I think</i> that sports have <i>no</i> borders.

Most themes of argumentative writing are general; everyone can discuss them without special knowledge. They are also controversial and disputable rather than neutral, so that there is no unanimous agreement but a variety of opinions and views. In that case, writers need to take their stance and explicitly state that it is their personal view. In Table 11, it is doubtful that everyone agrees that Japanese students are too lazy to study or that money is important. The death penalty is so controversial, provocative a theme that the writer took their own position with *I think* for his/her readers.

There is no obvious rubrics in NICE to require student writers to choose one of the two opposing views about themes (e.g., pros/cons or advantages/disadvantages), but

the learners were asked to state what they are going to write about. Intriguingly, the Japanese learners used *I think* to introduce their personal views, as seen in Table 12. The learner first framed the text, saying “I want to write three examples as below,” which corresponds to the NS correction, “I will give three examples of this.” After listing the three advantages of TV, the writer took a stance with *I think*, indicating what to follow is his/her own opinion rather than anyone else’s. *I think* also frames texts and guide readers by announcing one’s own opinion, differentiating it from others’ views. This explains why *I think* survived instead of being deleted (cf. Tang and John, 1999; McCrostie, 2008).

**Table 11. Surviving I think and general themes**

	The NNS writing	The NS correction
1	<i>I think</i> they are too lazy to study. [low JPN]	<i>I think</i> the students are too lazy to study.
2	<i>I think</i> money is very important. [low JPN]	<i>I think</i> money is very important.
3	<i>I think</i> we must justice the person who got the penalty by the same penalty. [middle JPN]	<i>I think</i> we must carry out the execution in a standard way.

**Table 12. Framing the texts**

The NNS writing	The NS correction
Although I said about good points, sometime these points may become bad influences for people. <i>I want to write three examples at below.</i> ...Third, because of TV, a conversation between families will be decrease. <i>I think</i> that it is important for us to think how to watch TV. For example when we watch TV, which programs we watch. [low JPN]	The things I have described as TV’s good points can sometimes lead to bad influences on people. <i>I will give three examples of this.</i> ... TV also causes conversation between family members to decrease. <i>I think</i> that we need to think about how we watch TV. For example, we should consider which programs we watch.

The JPN learners used *I think* as a frame marker to start ones’ own opinion, which was followed by the related statements, though their way of reasoning varied among the levels. JPN writers, in particular, the low JPN learners tended to roll up their argument, mostly based on personal experiences or perspectives, which are represented by the first person pronouns, as seen in:

(2) **I think** sports are very important. Not so hard exercise is good. But these days **I** don't sports. So **I** am gradually fat. **I** must sports. Sports make **us** healthy. Unless **you** sports, **your** muscles are gradually weak. So as soon as **we** active, **we** are tired. Our body is so weak. Sports make **us** happy. And encourage. Many athletes make **us** happy. And **we** are impressed. **We** are moved. But, in the sports world, it has very serious problems. In term play, some athletes think **we** must win. They have big pressures. So they are dangerous. They have some tricks. For example, they attack rival team member. Judge person doesn't notice. So not commit penalty. There live is very dangerous. **I think** they must have sportsman ship. **We** enjoy sports. No drug, no tricks. Now **I** try to sports, only 3 minutes. For on diet. Every day **I** keep exercise, **I** keep healthy. **I** encounter old man and lady every day jogging, walking. They continue it. It is very important, good habits. In sports, many people keep good relationship; enjoy talking, good time, and comfortable time. **I** want to spend good time in sporting. More exercise, more health, more happy. But too much exercise is not good and unhealthy. So be careful. And keep a good manner. [low JPN]

The low JPN writer first made a thesis statement, using *I think*, that is, "I think sports are very important," followed by the supporting sentence, "Not so hard exercise is good." However, after that, the writer stated his/her deviant, recent personal situations about sports, which fails to support the thesis statement. The second *I think* introduced another view of the writer's ("I think they must have sportsman ship"), followed by "We enjoy sports. No drug, no ticks." The point is that *I think* worked as a booster and a kind of frame marker even at the low JPN.

Interestingly, the following advanced learner personalized the theme at the beginning, too; the learner mentioned his/her preference for sports, referring to the personal experiences, as seen in the below:

(3) When **I** think of "Sports" the first words come up to my mind is "**I** don't like it". To tell the truth, **I** don't like sports especially because **I**'m not good at playing it. Until **I** graduated from high school, **I** had hard time spending "Physical Education" class. **I** had to do try many sports, but **I** was not good at it except swimming. However in these days, **I think** doing sports is important. There are two reasons why **I think** the sports is important. The first reason is that doing sports keeps **our** health. In these days, there are people who sit in front of the desk for a long time doing there duty including myself. If **we** sit in front of the desk for a long time, **we** feel sick especially **our** body; especially it is even hard to move because of the stress. So it is important to do sports, such as everyday exercise such as walking or stretch helps **our** body to keep healthy and relieve

the stress. The second reason is that doing sports also helps **our** mental health. **We** feel many stresses from the daily life or work, and sometimes they are hard to solve it even **we** think hard. When **we** feel bad, **we** feel **we** don't want to do anything. If **we** don't to anything and stay by ourselves, the situation won't change well. By doing sports, **we** can be together with somebody and won't think by ourselves. **I believe** that doing sports will help our health physically and mentally. However if **we** try new sports hard or start doing sports hard, sometimes they damage **our** body. **I think** it is important to start from easy training and make it harder if **we** feel good. Also it is important to enjoy playing sports.

Once the argument moved from his/her personal remarks or experience to their views, which was expressed by *I think*, the Japanese learner employed first person plurals or impersonal subjects instead of the first person singulars (e.g., *we feel ... / doing sports keeps our health ... / it is even hard ... / it is important ...*), which created involvement, objectivity, driving away subjectivity. This makes a sharp contrast with the low JPN. In (2), there were 10 occurrences of *I* including two of *I think* (NF per 1000, 45.87), while in (3) there were 11 instances of this pronoun comprising three of *I think* (NF per 1000, 36.54), suggesting that the low JPN packed more *I think* in the short passage.

Then, the writer of (3) concluded this passage by rephrasing his/her argument, using *I think* and *I believe*. This clearly indicates that *I think* was highly selective and was used in the same way as *I believe* in elaborating an argument and giving ones' own opinion (cf. Hyland, 2000, 2002). Importantly, *I think* as an alternative to *I believe* functions as a booster but the certainty would not be so high as other boosters such as *of course*, since it is a personal marker. To sum up, advanced writers did use this arguably problematic phrase in argumentation as a booster with certainty and a frame marker. It should be noted that general, daily topics of persuasive prose are open to personalization.

## Discussion

We have clarified the quality of *I think* in argumentation, which most of the previous studies of L2 writing have considered redundant and unnecessary in argumentation and argued that it should not be used in persuasive prose. *I think* is such a personal marker and functions as a hedge to lower the degree of certainty. As a result, it is not agreeable with emphatic expressions of high degree of certainty or probability. Using *I think* is also not appropriate in describing facts or situations to support one's own

views, thus, evidence, which requires objectivity. By the same token, this personal intruder cuts off a logical sequence like conditionals, consequences, and causes and effects. Writers should avoid the use of *I think* in such contexts. This quality of *I think* made NS proofreaders rewrite it into other phrases of certainty such as *should* and *seem* in our research.

The lack of the knowledge of the proper use of *I think* causes learners to choose it carelessly, where *I think* is inappropriate. From the deletion of the Japanese learners' *I think*, it has become clear that they were not so aware of the importance of logical relations or reasoning in argumentative writing or of distinguishing these relations from personalization with *I think*. The learners' misuse of *I think ... not* is not only a grammatical error but also a focal error: the NSs deleted it instead of rewriting it into *I do not think*, which would downgrade the persuasive writing due to the low degree of certainty.

On the other hand, the investigation of surviving *I think* confirms the fact that it also works as a booster in argumentative writing. The controversial themes require writers to choose to take a stance (i.e., pros or cons, or personal preference). General themes or topics, which do not require any expertise, enable writers to develop them on their personal experiences or episodes, triggering a high frequency of *I* including *I think*. In argumentation, *I think* is an essential tool to personalize themes so that writers can guide readers in their favor. *I think* marks a writer's personal, subjective opinion in argumentative writing, where other people may disagree or have another view. Especially in discussing controversial topics it is essential to distinguish one's own opinion from others. In fact, this marker is more frequently used in more formal argumentative writing as seen in applied linguistics and philosophy, which do not need scientific, hard evidence but manipulate concepts or elusive thoughts in a chain of reasoning. This is why Hyland (2000) put as this phrase a booster in his list of metadiscourse (cf. Tang & Johns, 1999).

It is not enough to tell that L2 writers should not use *I think* in argumentative writing or just reduce it. Some uses of *I think* by the Japanese learners were kept in the NS correction and were highly selective. As the essay by an advanced student showed, it is essential to take one's own position in arguing about controversial, provocative topics. In that case, *I think* is a common tool to announce one's view in argumentation. In fact, writers are required to take their stance in argumentative prose, as in TOEFL Writing Section and IELTS. Interestingly, in this investigation, the Japanese learners across the levels used this phrase to make thesis statement at the beginning without the explicit rubrics; they probably felt it is necessary to first state their stance about such controversial themes.

## Conclusion

Our method is unique in that parallel corpora, where learner English was used as a target and corrected essays by NSs as a source language to disclose the behavior of learners' *I think*, as compared with their appropriate forms by NSs on the line basis. This study shows that the Japanese writers had the overuse of *I think* against the NS correction; the Japanese writers used *I think* more frequently at the significant level than the NSs, which supports the view of the previous studies (Kobaysahi, 2009; McCrostie, 2008; Oi, 2002). While the previous research ended up with the quantitative results, our approach also revealed **how** learners employed this personal marker inappropriately in argumentation, referring to the corresponding sentences corrected by NSs. Despite the subjective quality of *I think*, the Japanese learners used it in supporting facts or with other boosters of the higher degree of certainty (e.g., *of course* and *probably*) than *I think*, and cut off the logical relations with it. On the other hand, the current research indicates that advanced writers did use *I think* in a right place of persuasive prose to tell readers their position about controversial topics, which has long been ignored in the studies of *I think*.

These findings renewed the previous research. Hyland (2000) did not distinguish *I think* from other boosters but actually the degree of certainty of *I think* is relatively low in comparison with other boosters. This research also suggest the relation between personalization and general topics, which McCrostie (2008) missed. It should be careful to say that the Japanese equivalent of *I think* as a hedging device may cause the overuse (Kobayashi, 2009). Not all of the categorical misuses by the Japanese learners are due to L1 transfer; they are lacking in the knowledge of the proper uses of *I think*, which ELT teachers should inform students of. From this we conclude that it is indispensable to raise L2 writers' awareness of the quality and the proper use of *I think* and the importance of a register of argumentative writing instead of summing up all the uses of *I think* as a discursive intruder in argumentative writing.

As demonstrated, the systematic, close comparison between the NNS writing and NS corrections are highly suggestive and beneficial for L2 learners. This research is applicable to classroom teaching; university students receive feedback about their common errors such as *I think* from the unique parallel corpora. This method is so direct, local and efficient that learners can see what their English should be. Actually, I used the parallel corpus for my writing class at university. I collected my students' writing individually through a web-based communication tool, and compiled the parallel corpus from their texts, creating the correspondence to my corrections. In class, the students' frequent errors from my own parallel corpus were displayed on

the students' monitors. In addition to the parallel corpus of my class I also used the results from the parallel corpus of NICE in this research in explaining elusive words or phrases like *I think*. It was efficient to explain common errors, using the parallel corpora in class, rather than to feedback each students and beneficial to add authentic corrections to my teaching of writing. The exact effectiveness of this teaching method should be investigated in the future research.

## References

- Ando, S. (2006) *TOEFL TEST taisaku iBT writing*. TOEFL [Preparation for TOEFL TEST, iBT writing]. Tokyo: Seminar Eigo Kyoiku Kenkyujo.
- Baker, P. (2004). Query keywords, questions of difference, frequency, and sense in keywords analysis. *Journal of English Linguistics*, 32(4), 346–359.
- Barlow, M. (2004). *ParaConc* [software].
- Biber, D., Johnsson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. London: Longman.
- De Cock, S., Granger, S., Leech, G., & McEnery, T. (1998). An automated approach to the phrasicon of EFL learners. In S. Granger (Ed.), *Learner English on computer* (pp. 67–79). London: Longman.
- Gillard, P., & Gadsby, A. (1998). Using a learners' corpus in compiling ELT dictionaries. In S. Granger (Ed.), *Learner English on computer* (pp. 159–171). Amsterdam/Philadelphia: John Benjamins.
- Granger, S. (1998). The computer learner corpus: versatile new source of data for SLA research. In S. Granger (Ed.), *Learner English on computer* (pp. 1–18). London: Longman.
- Granger, S. (2002). A bird's-eye view of computer learner corpus research. In S. Granger, J. Hung, & S. Petch-Tyson (Eds.), *Computer learner corpora, second language acquisition and foreign language teaching* (pp. 3–33). Amsterdam/Philadelphia: John Benjamins.
- Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge: Cambridge University Press.
- Hyland, K. (2000). *Disciplinary discourse*. Harlow, England: Pearson Education.
- Hyland, K. (2002). Authority and invisibility: authorial identity in academic writing. *Journal of Pragmatics*, 34, 1091–1112.
- Kobayashi, Y. (2009). Extracting characteristics of English essays written by Japanese EFL learners based on the NS/NNS text categorization model. *Proceedings of Jinmonkon Symposium Series of Information Processing Society of Japan*, 16, 261–268.
- Log-likelihood calculator of the university centre for computer corpus research on language. (n.d.). Retrieved February 7, 2010, from <http://ucrel.lancs.ac.uk/llwizard.html>.
- Lorenz, G. R. (1999). *Adjective intensification—Learners versus native speakers: A corpus study of argumentative writing*. Amsterdam/Atlanta: Rodopi.
- McCrostie, J. (2008). Writer visibility in EFL learner academic writing: A corpus-based study. *ICAME Journal*, 32, 97–114.

- Miki, N. (2008). Productive vocabulary in argumentative writing: The keyword analysis of TOEFL and LOCNESS. *Studies in English Language Teaching*, 31, 47–56.
- Miki, N. (2009). *The Influence of choice of reference corpora on the results of keyword analysis*. Unpublished MPhil Thesis, University of Birmingham.
- Oi, K. (2002). *Eigo mode de writing [Writing in the English mode]*. Tokyo: Kodansha.
- O’Keeffe, A., McCarthy, M., & Carter, R. (2007). *From corpus to classroom: Language use and language teaching*. Cambridge: Cambridge University Press.
- Pech-Tyson, S. (1998). Writer/reader visibility in EFL written discourse. In S. Granger (Ed.), *Learner English on computer* (pp. 106–118). London: Longman.
- Ringbom, H. (1998). Vocabulary frequencies in advanced learner English: a cross-linguistic approach. In S. Granger (Ed.) (1998). *Learner English on computer* (pp. 41–52). London: Longman.
- Sinclair, J. (2001). Preface. In M. Ghadessy, A. Henry, & R. L. Rosery (Eds.), *Small corpus studies and ELT: Theory and practice* (pp. vii–xiii). Amsterdam/Philadelphia: John Benjamins.
- Sugiura, M. (Ed.). (2008). *Eigogakushusha no collocation chishiki ni kansuru kisoteki kenkyu [A Fundamental study of English L2 learners’ collocational knowledge]*. Report of Grants-in-aid for Scientific Research (B) 2005–2007 (Project No. 17320084). Nagoya University.
- Tang, R., & John, S. (1999). The ‘I’ in identity: Exploring writer identity in student academic writing through the first person pronoun. *English for Specific Purposes*, 18, 23–39.