On the creation of a learner corpus for the purpose of error analysis

On the creation of a learner corpus for the purpose of error analysis

James Martin Rogers

Abstract

Learners with similar backgrounds have a tendency to make the same types of errors in L2 production. Such errors can be viewed as having the potential to inform pedagogical methodologies, in that they shed light onto which features of the L2 are the most problematic for particular learners. Analyzing such errors also provides insight as to why these learners tend to make these errors, thus furthering our understanding of how second languages are acquired.

This study aimed to create a learner corpus for the purpose of error analysis to discover which errors occurred most frequently, and to examine why such errors occurred. Various CALL (computer aided language learning) methodologies were utilized to create an approximately 85,000 word learner corpus. Errors were corrected and classified, and error analysis was conducted on the most frequent errors found. This analysis revealed that interference from the learners' L1 was the source for the majority of errors, while cultural and metalinguistic knowledge also proved to be at fault for some particular errors.

The results of this study should prove to be valuable for English language teachers and researchers in Japan, in that the most frequent English errors that Japanese learners produce were quantified and also discussed. Thus, teachers and researchers can be cognizant of which errors prove to be the most troublesome, and can better understand why they occur to help Japanese learners to avoid them.

Keywords: learner corpus, corpora, error analysis, CALL, Japanese learners

INTRODUCTION

When teaching a foreign language to groups of students with the same L1, teachers will begin to notice certain productive errors that occur again and again. Corder's (1967) learner error study revealed that many such errors are caused by L1 interference. The word 'error' inherently has a negative connotation, but many SLA researchers and teachers view

errors from a different vantage point. For example, Richards and Sampson (1974:4) believe that "errors should not be viewed as problems to be overcome, but rather as normal and inevitable features indicating the strategies that learners use." Others view errors as having the potential to inform our pedagogical approaches. For instance, Webb's (2006) book presents the most common 151 English mistakes he found that Japanese learners produce in an attempt to bring attention to errors that students have a high probability of producing. The ratio of such errors has yet to be revealed, though. So, the question of exactly which errors are the most common still remains.

Building upon such research, this study delves deeper into such common errors to quantify exactly which errors occur most frequently, and also examines why they occur. This is accomplished through the creation of a learner corpus and error analysis. The resulting data thus informs researchers and teachers in Japan as to the errors that are the most copious, thus providing guidance as to which warrant the most attention.

DATA COLLECTION

Tan (2005:126) states that "the goal of learner corpora research has been to gain better insights into learner language." Leech (1997) emphasized the value of such research, in that it provides authoritative answers for the specific kinds of errors learners produce. The field of learner corpus analysis has gained more and more attention over the years as technology has opened up new possibilities for extensive examinations of the types of errors language learners produce (Granger, 1998, 2002). Learner corpora can now be created more quickly and easily than before, and recently created corpora run from tens of thousands to hundreds of thousands of running words. For instance, Belz (2004) collected 46,000 running words of learner produced language for the purpose of analysis.

Subjects involved in this study were all Japanese university freshmen (n=165), and were from three separate bands on a 1-5 level banding system which banded students by Benesse Corporation's *GTEC* (Global Test of English Communication) (2004). Out of the total, 100 students were from level 1 (the highest level), 42 were from level 2, and 23 students were from level 5. Level 1 students' average score was 259.03, level 2 was 219.62, and level 5 was 164.83.

The goal of this research project was to create a 75,000-100,000 running word learner corpus from writing assignments over two semesters. Previous experience with the target

learners led to the realization that motivation to do homework can occasionally be low, and lead to copying, doing the bare minimum, the use of online translation, plagiarism, etc. Researchers such as Crookes and Schmidt (1991) and Dornyei (2001) have found that motivation is crucial for language learning and that teachers can have an impact on students' motivation levels. Thus, a considerable amount of effort was made to make the writing assignments enjoyable and easy to submit by using multiple CALL (computer aided language learning) methodologies. A website was created for this purpose, which consisted of 10 separate homework assignments for each class. The goal of the assignment was to read a passage (reading and writing classes) or listen to an audio file (speaking and listening classes) on one of the 10 webpages, and write a reaction to it. Each passage contained approximately 250 words. Nation (2001) states that for adequate comprehension to occur, 95-100% of the running words in a text need to be known and should be high frequency vocabulary. Therefore, each passage contains no more than 5 words (98%) that had a high probability of being unknown to the students. In total, 40 separate passages were created, with 10 being assigned to each of 4 groups.

Students were instructed to write a minimum of 50 words in reaction to the passage or audio file. To motivate students to write more, and thus have a larger, more reliable corpus, photographs and videos were embedded into each webpage that were connected to the passage's theme. Each assignment consisted of one passage and from 1-5 photographs and 1-5 videos. Students were instructed that they must react to the passage, but as an option, they could write their reaction to the photographs/videos as well. Each passage was created with the intent to pique the students' interest, and thus the topics, photographs, and videos are humorous, shocking, interesting, etc. For example, one webpage's topic is a famous band whose members actually attended the same university as the students themselves. Another webpage contains a video of a shocking news report about a planned skyscraper which will have floors which revolve on their own axis. Each passage has a different topic with the goal of collecting a wide variety of errors.

A considerable amount of effort was also made to make the assignments easy to access and submit. Each assignment is on one single webpage which scrolls, so that students do not have to click on various pages to view videos, photos, etc. In addition, custom html code was embedded into the page to enable students to submit their reaction to the webpage without going through the hassle of opening Microsoft Word, typing their reaction in it, checking the word count, saving the file, logging in to their email, creating an email, attaching their

assignment, and finally sending the file. The code creates a text box which students can type their reaction, their data (name, homework assignment #, class #, and email), and also contains a word count function which automatically counts the words as the students type. Each text box for inputting students' email automatically uploads '@kansaigaidai.ac.jp', since this is the ending of each student's university email. Thus, all students have to add is their 7 digit student number. This feature was added after a large amount of students misspelled their own email address over and over. For example, students would type '@knsaigaidai. ac.jp' or '@kansaigaidai.co.jp'. So, researchers using internet technologies need to pay heed to even simple tasks like this which most would assume students wouldn't have a problem with. These features make completing an assignment simple and quick, thus avoiding the confusion and frustration that commonly accompanies computer technology. By clicking send in the text box, an email is automatically sent to the teacher. When the email is received by the teacher, he/she can send corrections to students seamlessly without having to retype students' emails simply by clicking reply. The system automatically attaches markers for each section as well, which enables a teacher to copy and paste all students' data at the end of a semester from each email into Microsoft Excel, and then sort. For example, in each email the marker 'myMessage =' is attached to the beginning of each student's reaction along with a page break, so a quick sort can easily sort all messages together. Similarly, the marker 'wordCount =' is attached to each word count result for each assignment, allowing for teachers to easily sort and calculate average word counts by students.

ERROR ANALYSIS

This study utilizes Lennon's (1991:182) definition of 'errors', which he defines as "a linguistic form or combination of forms which, in the same context and under similar conditions of production, would, in all likelihood, not be produced by the speakers' native speaker counterparts." Errors will be classified according to Dulay, Burt, and Krashen's (1982) taxonomy for describing learner errors, which breaks them down into the following categories: omission, addition, misinformation, and misordering. Spelling mistakes are ignored since they are automatically highlighted when typing on the website, and possible corrections are given when you right click. Thus, counting such errors would be unreliable since many more such errors would occur without the aid of web based spell correction. Listed below are the types within these 4 categories.

1. Omission (Article, Lexical, Number, Possessive, Preposition, Pronoun, Run-on Sentence)
2. Addition (Article, Lexical, Number, Possessive, Preposition, Pronoun, Sentence Fragment)
3. Misinformation (Article, Lexical, Possessive, Preposition, Pronoun, Verb Tense, Word Form)
4. Misordering (Word Order)

RESULTS

Overall, the collection method and resulting quantity and quality of the data in the learner corpus was successful. On average, students produced 72.6 words per homework, 45.2% more than the minimum requirement of 50 words. Students tended to write more when the topic of the webpage interested them. The highest word count for any particular homework was 469. Many students also commented on post course evaluations that they enjoyed doing the homework and that the webpages were interesting. Overall, students did 85.2% of the 10 homework assignments. These factors led to a learner corpus of 85,246 running words being created.

Except for the lowest level class, the quality of the writing was ideal. There were no instances of students copying from each other or plagiarizing, and only 1 student had to be warned not to use an online translator. The level 5 class' data had to be excluded from this study, though. The task of free writing in reaction to the websites was too challenging for them, resulting in very heavy use of dictionary translations. Students were not cheating per se (typing entire reactions in Japanese into online translators), but rather were taking phrase after phrase from translations and splicing them together, resulting in quite broken English strewn with errors. Unfortunately, the task may have just been beyond their capacities.

In total, 3661 distinct errors were corrected. The top 10 error types corrected accounted for 82.6% of all student errors. The most frequent error was lexical based on misinformation (i.e., the wrong lexical item was used). This error accounted for 23.5% of all errors. The following are the top 10 error categories, types, and number of occurrences: Misinformation (Lexical): 861 errors, Omission (Article): 560, Omission (Number): 420, Omission (Lexical): 308, Misinformation (Word Form): 213, Misinformation (Verb Tense): 200, Omission (Preposition): 187, Misordering (Word Order): 158, Addition (Lexical): 118, and Misinformation (Preposition): 113.

Upon further analysis, breaking down the types into categories reveals that the most frequent specific type of error is actually Omission (Article) with 560 occurrences. So,

although Misinformation (Lexical) had many more occurrences (861), the variety within this category brings its ranking down when the data is examined for specific errors. However, if Omission (Article: a) and Omission (Article: the) are to be considered as distinct errors, then Omission (Number) becomes the most frequent distinct error with 420 instances. For a more detailed list of the most common error types and their frequencies, please see Figure 1. It lists the specific error types or the top 3 examples of specific error types found for types which have variety.

When the results of this study are compared to Webb's (2006) *151 Common Mistakes of Japanese Students of English*, striking similarities are found. The vast majority of the common mistakes illustrated in his book are present in the corpus, save for a handful. For example, the common mistake of writing 'pierce' instead of 'earring' was not present, but this does not indicate that the mistake is not common. Rather, errors are highly context sensitive, so if one of the topics of the website passages was connected to jewelry or fashion, such an error may just as well have cropped up. These finding indicate that the passages and their topics were wide enough to extract a large variety of errors, similar to the ones found by Webb (2006). There were some frequent errors that were not listed in his book, but again, this is indicative more of contextual influence rather than any kind of shortcoming of Webb's (2006) book. They are:

1. 'video': Learners often refer to internet videos as 'animations', 'moving images', or 'movies'

2. 'drive a car': Learners often write 'ride a car' or 'take a car'

3. 'look at a picture': Learners often write 'watch a picture'

4. 'picture': Learners often refer to pictures on a website as 'images'

5. 'envy': Learners often write 'envy' instead of 'jealous' or a more appropriate phrase

6. 'story/passage': Learners refer to the story in the passages on the webpages as follows: 'I heard your speech' or 'I heard your speaking'

7. 'say/said': Learners often write 'talk' instead of 'say/said', such as in 'I can't understand what batman talks'

8. 'Mr.': Learners often attach 'Mr.' to first names, although they are cognizant of the fact that it is the person's first name

9. 'guy': Learners often refer to a young man as 'man' instead of the more appropriate 'guy'

10. 'foreign': Learners often use 'abroad' instead of 'foreign', as in 'don't know abroad dramas'

DISCUSSION

L1 interference can explain many types of L2 errors, and has been cited by various researchers as a major obstacle in obtaining proficiency in an L2. For example, Areas Da Luz Fontes and Schwartz (2010), Elston-Guttler, Paulmann, and Kotz (2005), and Schwartz, Yeh, and Shaw (2008) all found the lack of one-to-one correspondence in meaning between L1 and L2 to be the source of learner errors. This study confirmed previous research's findings that L1 is a major source of learner errors, in addition to revealing other error sources, such as simple lack of knowledge and also cultural and conceptual influence. The following is a discussion of how and why the top 10 error types found in this study occur.

Lexical errors were the most numerous in this study, with Omission, Addition, and Misinformation all being in the top 10 errors. Misinformation (Lexical) was the most common error in the entire corpus. When these errors were examined, one particular type of error occurred quite frequently. This error seemed to stem from over-reliance on *denshi jisho*, or electronic dictionaries. This overreliance on these dictionaries, or what this author deems as *denshi jisho syndrome*, can be quite detrimental to a learner's writing and nullify any benefits that the learner gains from their use. For example, one learner produced 'he will certainly form a lump', referring to someone who had gotten hurt. It is assumed that the phrase 'form a lump' came from an electronic dictionary, in that the likelihood that the learner was taught such a low frequency phrase during their high school years, or that it was learned on their own is quite low. Inappropriate or odd phrases such as 'form a lump' cropped up again and again in the corpus, and the likely culprit is electronic dictionary over-reliance. Further research, such as a follow-up study and an in-depth examination of the types of language electronic dictionaries produce, should be conducted to verify these findings.

Kobayashi (2006) and Chon (2008) both note the rise of popularity of electronic dictionary usage in recent years, particularly with Asian learners. Both researchers also acknowledge their shortcomings. Kobayashi (2006:iii) remarks that an "increase in the frequency of dictionary consultation may be accompanied by varying degrees of decrease in the frequency of inferring" for some Japanese students. Chon (2008:23) also noted similar problems with Korean students' writing, in that such aids "led writers to identify further dictionary-based lexical problems due to writers' lack of knowledge of the L2 word(s) offered." Both Kobayashi (2006) and Chon (2008) recommend training students to properly use such aids.

Not all lexical errors stemmed from electronic dictionary use, though. Many also stem

from the learner's L1. For instance, a common Omission (Lexical) error written by multiple learners was referring to 'family members' as 'family'. When reacting to one passage about someone whose family member died, one learner wrote 'he lost his family' instead of 'he lost his family member'. This is clearly influenced by the Japanese, which refers to the death of even one family member simply as 'family'. The above student probably directly translated 'ご家族がなくなった' directly into English, and thus ended up with the error 'he lost his family.' A similar L1 influence can be seen in multiple Addition (Lexical) errors as well, such as with the simple error 'James teacher, too'. Students are obviously attaching '先生' to their teacher's name, regardless of the fact that they were all aware that James is their teacher's first name.

This study also found that the articles of English account for a large percentage of learner errors. A number of studies have examined why articles seem to be so difficult for specific language learners. For example, studies have found that when a learner's L1 lacks articles, they have more trouble with an L2 which utilizes articles in comparison with learners whose L1 utilizes them (Ionin, Ko, Wexler, 2004; Hawkins, et. al, 2006). Butler (2002:472) found that depending on the proficiency level, the following hypotheses are utilized by learners:

1. context-insensitive hypotheses
2. hypotheses that showed sensitivity to inappropriate contextual cues
3. hypotheses that showed sensitivity to a range of relevant contexts

The results of these studies in addition to the current study highlight the difficulty that learners have with the articles of English, and justifies extended classroom time dealing with these issues. Omission (Article) was the only article type within the top 10 most common errors, but it is important to note that Addition (Article) and Misinformation (Article) also occurred in significant numbers. These results reveal that the learners had difficulty with all facets of article usage. Bulter (2002:476) suggests that "article instruction would be more effective if it were incorporated into various exercises in countability detection wherein the students experience how native speakers change their perception of an entity depending on the context." Hiki (1990) also agrees that countability detection is central to proper article usage.

Studies have shown the lack of obligatory plural marking in Japanese lead Japanese learners to have difficulty with the English plural (Wakabayashi, 1998; Hakuta, 1978). This is obviously evident in that Omission (Number) was the third most common error found. For

example, one learner wrote 'I'm interested in taking picture'. There is clear L1 influence in that the Japanese word for 'picture' (写 真) does not change form regardless of if it is referring to the singular or plural. Learners from other languages that lack plural markings, such as Chinese, have also been shown to have difficulty with this feature of English (Jia, 2003). Interestingly, Athanasopoulos and Kasai (2008) found Japanese learners who have problems with English plurals tend to think of English nouns in terms of their material qualities and not their shape, as native English speakers do. The problem thus not only stems from language features, but also how individuals conceptualize language. Young-Davy (2000) and Butler (2002) also found such a two-fold issue, where not only language features, but also metalinguistics played a role in error production. The fifth most common error in this study was Misinformation (Word Form). These kinds of errors may be influenced by L1 interference like many of the other errors in the top 10, but more often seem to stem from a lack of knowledge of how morphological affix changes should be employed. On one hand, some research has pointed to Japanese learners' L1 having some influence on prefix knowledge. Mochizuki and Aizawa (2000) note that common loanwords in Japanese contain very common English prefixes, such as 're' in the loanword 'recycle', 'un' in 'unlucky', 'non' in 'nonstop', and 'anti' in 'anti-aging' and that knowledge of these loanwords lead learners to mastery of the prefixes they contain. This is noteworthy in that none of the word form errors were with prefixes, but rather all were with suffixes. Previous research in regards to Japanese students' knowledge of English suffixes has shown that it is quite poor, even by high school or university. For instance, Schmitt and Meara (1997) found Japanese learners' knowledge of allowable suffixes for verbs, especially derivative suffixes, to be quite poor. While the results of this study correspond with the finding that suffix knowledge is poor, an examination of whether suffix errors were derivational or inflectional resulted in data that conflicted with Schmitt and Meara's (1997) finding. Inflectional suffix errors accounted for a much larger percentage of the total Misinformation (Word Form) errors at 65.3%, compared to only 34.7 % being derivational suffix errors. Therefore, Japanese learners' knowledge of suffixes, and in particular inflectional suffixes, seems to be the most deficient. More research should be done to resolve the discrepancy between this study's findings and Schmitt and Meara's (1997) finding.

English verb tense also proved difficult for Japanese learners, and studies have also found L1 to be the culprit. In her study of Japanese and French learners, Collins (2007:302) found that "an ESL/EFL learner's L1 may indeed result in the formulation of

inappropriate hypotheses about how tense-aspect forms work in English." This is evident in the misunderstanding of the phrasal verb 'get married' in errors found such as 'I want to get marry.' To address this issue, Collins (2007:300) found that "instructional activities that require learners to supply or manipulate contexts for given forms may help address the difficulty learners have." The L1 can cause interference in different ways as well. For example, since the Japanese loanword 'リラクス' (relax) does not change tense to express sentences such as 'he looked very relaxed', it is not surprising that one learner made the error 'he looked very relax' in the corpus.

What is not evident, though, is why errors that were clearly not influenced by the L1 occurred. The vast majority of the most common verb tense error (using present tense instead of past) occurred despite the fact that the Japanese translations of the proper English contained past tense verbs. For example, one learner writes 'When I was a high school student, after school I often go to movie with my friends.' The learner clearly exhibits mastery of past tense with other verbs in the sentence, but for some reason opts for the present tense of the verb 'go', despite the fact that a translation would yield the past tense for go (行った). Such errors were by far the most numerous, and it is unclear why they occurred. The errors seem careless, but why would learners who were careless write on average 45.2% more than necessary? The free nature of the writing assignments and the focus on making assignments fun and motivating may actually increase the number of careless errors. While this is only speculation, this answer seems most logical. Still, more research needs to be done to figure out the true source of such errors.

The seventh most common error in this study was Omission (Preposition), with the top three particular omitted prepositions being 'to', 'in', and 'of'. At first glance, these kinds of errors seem peculiar since the learners' L1 utilizes prepositions, albeit in different ways depending on the context. So, why are these learners omitting prepositions when an L1 translation would use them and all previous education should lead students to realize that a preposition is needed? The answer can be found when we examine the errors of the 11th most common error found in this study: Misinformation (Preposition). The following are the top three such errors and sample example sentences from the learner corpus:

1. in/on: 'went there in school trip' instead of 'went there on a school trip'

2. of/in: 'the man of the first video' instead of 'the man in the first'

3. on/in: 'only on pictures' instead of 'only in pictures'

Why do students misuse prepositions in such ways? Young-Davy (2000:iv) found that

"English and Japanese speakers do not conceive of spatial relationships in the same way." This leads to different prepositions being used in different contexts. So, for example, in the above three example sentences, L1 influence is clear in that there is an overgeneralization strategy being employed. In the first example, 'in' is utilized in place of 'on'. When the sentence is translated into Japanese, the preposition に is used. It is thus assumed that learners are tapping previous knowledge of に usage which often translates into 'in'. This can be seen in how the phrase '部屋にあるテレビ' translates into 'the TV in the room', and thus this error occurs.

So, what exactly is the connection between Misinformation (Preposition) and Omission (Preposition) errors? Many teachers and researchers have noted the cultural different between Western culture and Japanese culture in regards to modesty. Loveday (1982) remarks that the Japanese are hesitant to speak out. Prichard (2006:153) notes that Japanese learners "display self-deprecation in underestimating their own ability." So, why do Japanese have a tendency to be so reserved? Berwick and Ross' (1989) study on Japanese learners' motivation found that they have difficulty overcoming their dependence on teachers and authority figures. Takahashi (1999:5) agrees, in that he believes Japanese learners' reluctance to speak out stems from two primary reasons: "the Japanese education system and the unique Japanese culture."

Therefore, the only plausible reason for such simple preposition omission errors is that the learners have become cognizant of the differences in conceiving spatial relationships that Young-Davy (2000) noted, but are still unsure as to which preposition is appropriate. This uncertainty is possibly prompting learners to abstain from taking the risk of making an error by adding a preposition they are unsure of, and thus they opt to omit it. Culturally, such an action may be the safest choice in their minds.

In regards to the ordering of words, Chang (2009:374) states that "particular languages have particular biases for ordering words or phrases." This is an issue for the Japanese learner of English, as Izzo (2000:133) points out that such errors "likely result from grammatical and structural differences between English and their native language." Many researchers who examine word order differences between English and Japanese concentrate on the structural difference of English's SVO order and Japanese's SOV order (Sawasaki, 2007; Wang, Gong, and Minnett, 2009). Such errors were evident in this corpus, as tendency for Japanese learners to put verbs last in sentence order can be seen in 'He is a good singer I think.' 'I think' stems from 'と思います,' which is typically placed last in such a sentence in

Japanese.

This study found that micro-ordering conventions were transferred from the L1 as well. For example, multiple learners put 'I' before other nouns and pronouns, such as with 'I and my family.' This error is influenced by the L1 '私と私の家族' where 'I' typically comes before 'family.' Typically, the subject is not included when such phrases are translated into Japanese, but when it is, the word order sometimes follows the English convention, while at other times does not. Another common error was the placement of adverbs before the subject in a sentence. Many learners put adverbs such as 'probably' and 'especially' before the subject. For example, '私は多分' (I probably) and '多分私は' (Probably I) are both possibilities in Japanese. It may be the case that occasionally students are directly translating the latter, which is not the typical grammatical convention in English.

CONCLUSION

This study created a learner corpus and conducted error analysis on the errors students produced. Considerable effort was put into creating assignments which would yield the ideal learner corpus in terms of quality and quantity. The resulting 85,246 word corpus was deemed a success on many facets. It proved to be easy to collect, and student motivation to write remained high throughout, resulting good quality and also resulted in learners writing 45.2 % more than what they were required. Although, one fault of the study was that the assignments proved to be too difficult for level 5 students.

The errors found in this corpus were corrected and classified. Thus, this study was able to quantify the errors Japanese students produce, and provide data as to which were the most common. The top 10 most common errors were then broken down by type, and analyzed for reasons why they occurred. This error analysis revealed that the learners' L1 was at fault for a vast majority of the errors found, but also that metalinguistic knowledge plays a role in error production. Error analysis conducted in this study does warrant further research to be done to validate assumptions made on why learners produced such errors.

Hopefully, the data provided by this study will help guide teachers in Japan in materials creation and selection. For example, the results of this study supported the selections made by Webb (2006), and thus his book would undoubtedly be an asset for any Japanese university freshman. Also, the data about specific errors should also prove useful for teachers or materials writers in that it provides information as to which features of English they

should bring attention to, thus helping students avoid errors they have a high probability of producing. Finally, it is hoped that researchers will find value in the methodology that this study employed to create a learner corpus and also in the error analysis conducted.

Figure 1: The most frequent specific error types

#	Error category, type, and specific error and/or example where applicable
420	Omission (Number): such as in 'a lot of mistake'
103	Misinformation (Lexical): students used 'movie' instead of 'video'
91	Misinformation (Verb Tense): students used present instead of past tense
79	Addition (Number): such as in 'are kind for families'
65	Addition (Article): 'the', such as in 'a way of the communication'
59	Omission (Preposition): 'to', such as in 'they are kind each other'
49	Addition (Article): 'a', such as in 'becoming a *geisha* is a very hard work.'
49	Omission (Article): 'a', such as in 'but he became rich man'
43	Omission (Preposition): 'in', such as in 'I knew that *sanshin* was used Okinawa'
42	Misinformation (Article): students used 'a' instead of 'the' and vice-versa, such as in 'you are a happiest guy'
31	Addition (Preposition): 'to', such as in 'I have never been to there'
30	Misinformation (Lexical): students used 'envy' instead of 'jealous', etc.
28	Omission (Pronoun): 'it', such as in 'I tried to do but I'
27	Misinformation (Verb Tense): students used past instead of present tense
27	Omission (Lexical): 'am', such as in 'I surprised'
27	Omission (Lexical): 'is', such as in 'your wife so beautiful'
24	Omission (Preposition): 'of', such as in 'take photos foreigners'
16	Misinformation (Article): students used 'a' instead of 'an' and vice-versa
16	Misinformation (Possessive): 'of', such as in 'I have a nice lens of Leica'
15	Omission (Lexical): 'was', such as in 'what he doing'
13	Addition (Preposition): 'in', such as in 'I want to drive somewhere in this summer'
13	Misinformation (Lexical): students used 'ride' instead of 'drive'
13	Misinformation (Preposition): students used 'in' instead of 'on'
12	Misinformation (Verb Tense): students used present instead of progressive tense
12	Omission (Pronoun): my, such as in 'I don't have own car yet'
11	Misinformation (Preposition): students used 'of' instead of 'in'
11	Omission (Pronoun): 'I', such as in 'the country that wants to go is Spain'
10	Addition (Sentence fragment): such as in 'I am interested in this story. Because'
10	Misinformation (Preposition): students used 'on' instead of 'in'
9	Addition (Preposition): 'of', such as in 'I like square of cars'

8	Misinformation (Possessive): students made mistakes such as 'James's'
8	Omission (Possessive): such as in 'my father side of family'
7	Addition (Lexical): 'is', such as in 'your family is looks very happy'
7	Misinformation (Pronoun): students used 'it' instead of 'them'
7	Misinformation (Word Form): 'scary', such as in 'I feel scary when I watch…'
7	Misordering (Word Order): 'I think', such as in 'He is a good singer I think.'
7	Misordering (Word Order): pronoun + pronoun, such as in 'I and my family'
7	Omission (Lexical): 'and', such as in 'I will go to Kyoto tomorrow, I hope to…'
6	Addition (Lexical): 'do', such as in 'I can't do wink well'
6	Addition (Lexical): 'like', such as in 'the monkey looks like very comfortable'
6	Addition (Lexical): 'was', such as in 'I was laughed at this video'
6	Misinformation (Pronoun): students use a proper noun instead of 'you'
6	Misordering (Word Order): 'especially', such as in 'Especially I don't like…'
5	Addition (Lexical): 'and', such as in '…meet the man and talk to man and I…'
5	Misinformation (Pronoun): students used 'it' instead of 'I'
4	Addition (Possessive): such as in 'I don't know what monkey's said'
4	Addition (Pronoun): 'my', such as in 'If I have my children'
4	Misinformation (Word Form): 'surprise', such as in 'It was surprised that'
2	Addition (Pronoun): 'one', such as in 'I couldn't see the last one picture'
2	Addition (Pronoun): 'them', such as in 'I could understand them by the video'
2	Omission (Run-on sentence), such as in 'to become a good photographer all the more there might be'

WORKS CITED

Athanasopoulos, P., & Kasai, C. (2008). Language and thought in bilinguals: The case of grammatical number and nonverbal classification preferences . *Applied Psycholinguistics, 29*, 105–123.

Areas Da Luz Fontes, A., & Schwartz, A. (2010). On a different plane: Cross-language effects on the conceptual representations of within-language homonyms. *Language and Cognitive Processes, 25*(4), 508-532.

Belz, J. (2004). Learner corpus analysis and the development of foreign language proficiency. *System, 34*(4), 577-591.

Benesse Corporation (2004). *GTEC* (Global Test of English Communication). Retrieved from http://www.benesse.co.jp/gtec/english/index.html

Berwick, R., & Ross, S. (1989). Motivation after matriculation: Are Japanese learners of English still alive after examination hell? *JALT Journal, 11*, 193-210.

Butler, Y. (2002). Second language learners' theories of the use of English articles: An analysis of the

metalinguistic knowledge used by Japanese students in acquiring the English article system. *SSLA, 24*, 451-480.

Chang, F. (2009). Learning to order words: A connectionist model of heavy NP shift and accessibility effects in Japanese and English. *Journal of Memory and Language, 61*, 374-397.

Chon, Y.V. (2008). The electronic dictionary for writing: A solution or a problem? *International Journal of Lexicography, 15*(3), 206-228.

Collins, L. (2007). L1 differences and L2 similarities: teaching verb tenses in English. *ELT Journal, 61*(4), 295-303.

Corder, P. (1967). Significance of learner's errors. *IRAL 5*, 151-169.

Crookes, G. & Schmidt, R.W. (1991). Motivation: reopening the research agenda. *Language Learning, 41*, 469–512.

Dörnyei, Z. (2001). *Teaching and researching motivation*. Harlow: Longman.

Duly, H., Burt, M., & Krashen, S. (1982). *Language Two*. Oxford: Oxford University Press.

Elston-Güttler K E, Paulmann S, Kotz S A. (2005). Who's in control? Proficiency and L1 influence on L2 processing. *Journal of Cognitive Neuroscience, 17*(10), 1593-1610.

Granger, S. (1998). *Learner English on Computer*. London: Longman.

Granger, S., Hung, J. and Petch-Tyson, S. (2002). Computer learner corpora. *Second Language Acquisition and Foreign Language Teaching*. Amsterdam: John Benjamins.

Hakuta, K. (1978). A report on the development of the grammatical morphemes in a Japanese girl learning English as a second language . In E. Hatch (Ed.), *Second language acquisition* (pp. 132–147). Rowley, MA: Newbury House.

Hawkins, S., et al., (2006). Accounting for English article interpretation by L2 speakers. *EUROSLA Yearbook, 6*, 7-25.

Hiki, M. (1990). The judgment of noun countability by Japanese college students. Where is the difficulty? *JACET Bulletin, 21*, 19-38.

Ionin, T., Ko, H. & Wexler, K. (2004). Article semantics in L2 acquisition: the role of specificity. *Language Acquisition, 12*, 3-69.

Izzo, J. (2000). Sentence structure aberrations in English writings of Japanese university students. In T. Orr & K. Cranker (Eds.), *Center for Language Research 1999 Annual Review* (pp. 133-140). Aizuwakamatsu, Japan: University of Aizu.

Jia, G. (2003). The acquisition of the English plural morpheme by native Mandarin Chinese speaking children. *Journal of Speech, Language, and Hearing Research, 46*, 1297–1311.

Kobayashi, C. (2006). *The use of pocket electronic dictionaries as compared with printed dictionaries by Japanese learners of English* (Unpublished doctoral dissertation). Ohio State University, Columbus, OH.

Leech, G. (1997). Teaching and language corpora: a convergence in A. Wichmann, S. Fligelstone, T. McEnery, and G. Knowles (Eds.). *Teaching and Language Corpora*. Harlow: Addison Wesley Longman.

Lennon, P. (1991). Error: Some problems of definition, identification, and distinction. *Applied Linguistics, 12*(1), 180-195.

Loveday, L. (1982). Communicative interference: A framework for contrastively analyzing L2 communicative competence exemplified with the linguistic behaviour of Japanese performing in English. *International Review of Applied Linguistics, 20*, 1-16.

Mochizuki, M., & Aizawa, K.(2000). An affix acquisition order for EFL learners: an exploratory study. *System, 28*, 291-304.

Nation, I.S.P. (2001). Learning vocabulary in another language. Cambridge: Cambridge University Press.

Pritchard, R. (2006). The changing self-perceptions of Japanese university students of English. *Journal of Studies in International Education, 10*, 141-156.

Schmitt, N. & Meara, P. (1997). Researching vocabulary through a word knowledge framework. *SSLA, 20*, 17-36.

Schwartz, A.I., Yeh, L., & Shaw, M. (2008). Lexical representation of second language words: Implications for second language vocabulary acquisition and use. *Journal of the Mental Lexicon, 3*, 309-324.

Sawasaki, K. (2007). *L2 reading by learners of Japanese: A comparison of different L1s* (Unpublished doctoral dissertation). Ohio State University, Columbus, OH.

Takahashi, K. (1999). Why are Japanese students so quiet in America? *Northwest Asian Weekly, 18*(40), 5.

Tan, M. (2005) Authentic language or language errors? Lessons from a learner corpus. *ELT Journal, 59*(2), 126-134.

Wakabayashi, S. (1998). Systematicity in the use of the definite article by Japanese learners of English. *Gunma Women's University Proceedings, 19*, 91-107.

Wang, W., Gong, T. & Minnett, J. (2009). A simulation study on word order bias. *Interaction Studies, 10*(1), 51-75.

Webb, J. (2006). 151 *Common Mistakes of Japanese Students of English*. Tokyo: The Japan Times.

Young-Davy, B. (2000). *A cognitive-semantic approach to the acquisition of English prepositions* (Unpublished doctoral dissertation). University of Oregon, Eugene, OR.

（James Martin Rogers　国際言語学部講師）