Using Learner Corpora for L2 Lexicography: Information on Collocational Errors for EFL learners

Yukio Tono, *Department of English Education, Tokyo Gakugei University, Tokyo, Japan*

Abstract: In this paper, we describe an on-going project of the corpus of EFL (English as a Foreign Language) learners in Japan and its application for pedagogical dictionary compilation. We especially focus on the learners' errors in verb collocation patterns and describe how the learner's dictionary can benefit from the learners' error information based upon the learner corpora.

Keywords: CORPUS, LEARNER CORPORA, CORPUS LINGUISTICS, PEDAGOGICAL DICTIONARY, L2 LEXICOGRAPHY, LEARNER'S DICTIONARY, APPLIED LINGUISTICS, COLLOCATIONS, ERROR ANALYSIS, COLLOCATION ERRORS, VERB PATTERNS, SECOND LANGUAGE ACQUISITION RESEARCH, DICTIONARY USER INFORMATION, USAGE COLUMN, NEGATIVE INFORMATION

Opsomming: Gebruik van aanleerderskorpusse vir L2-leksikografie: Inligting oor kollokasionele foute vir EVT-aanleerders. In hierdie artikel beskryf ons 'n voortgaande projek oor die korpus van EVT- (Engels as 'n vreemde taal-) aanleerders in Japan en die toepassing daarvan vir die samestelling van pedagogiese woordeboeke. Ons fokus veral op die aanleerdersfoute gemaak ten opsigte van werkwoordkollokasionele patrone, en beskryf hoe die aanleerderswoordeboek voordeel kan trek uit die aanleerders se foutinligting gebaseer op die aanleerderskorpusse.

Sleutelwoorde: KORPUS, AANLEERDERSKORPUSSE, KORPUSLINGUISTIEK, PEDAGOGIESE WOORDEBOEK, L2-LEKSIKOGRAFIE, AANLEERDERSWOORDEBOEK, TOEGEPASTE LINGUISTIEK, KOLLOKASIES, FOUTANALISE, KOLLOKASIONELE FOUTE, WERKWOORDELIKE PATRONE, NAVORSING OOR TWEEDETAALVERWERWING, WOORDEBOEKGEBRUIKERSINLIGTING, GEBRUIKSKOLOM, NEGATIEWE INLIGTING

1. Introduction

Recent development of corpus linguistics and actual corpora has been remarkable. Many dictionaries published recently all enjoyed in some way or another the use of large corpora; for example, the *COBUILD English Dictionary* used the Bank of English, the corpus of 20 million words in contemporary English,

developed at the Birmingham University. *Longman Dictionary of Contemporary English* and *Oxford Advanced Learners' Dictionary of Current English* used the British National Corpus, produced by an academic and industrial consortium consisting of Oxford University Press, Longman, Chambers Harrap, Oxford University Computing Services, Lancaster University's Unit for Computer Research on the English Language and the British Library. Older corpora for research were all gathered by the institution ICAME (International Computer Archive of Modern English), which is an international organization of linguists and information scientists working with English machine-readable texts. The aim of the organization is to collect and distribute information about English language material available for computer processing, and about linguistic research on this material, completed or in progress, in order to compile an archive of English text corpora in machine-readable form, and to make material available to research institutions.

Even though many different kinds of corpora have been available internationally, very few researchers have yet built up a corpus of the language learner. There may be a couple of reasons for this; first, the data collected from language learners is in most cases erroneous. The primary interest of the corpus builders at present is to describe the status quo of native speakers' language, so they are basically not interested in collecting learner language data. Secondly, and related to the first, most of the researchers in applied linguistics or TESL (Teaching English as a Second Language) / TEFL (Teaching English as a Foreign Language) are not sufficiently informed about the expertise of corpus linguistics. They are either more or less classroom-oriented researchers or theoreticians like UG-based SLA researchers (Universal Grammar-based Second Language Acquisition researchers), who stress the intuition of the native speaker, rather than the collection of a large text.

More and more attention has been paid, however, to building a corpus of the language learner. To date, International Corpus of Learner English (ICLE) has been one of the largest and the most systematic corpus development projects in the world (Granger 1994). Longman has been developing Learners' Corpus for its dictionary project[1]. John Milton of Hong Kong University of Science and Technology has already collected about 8 million words of the writing of Chinese students of English[2]. In Japan, Asao and others held a symposium on EFL learner corpora and SLA (Asao et al. 1995). Our project is also one of the few attempts to develop learner corpora.

In this paper, we will first describe our learner corpus project and then show the application of learner corpora data to English pedagogical dictionary-making as an example and examine the potential which learner corpora have for future L2 lexicography.

2. TGU Learner Corpus Project

Tokyo Gakugei University, a national teacher training college in Tokyo, Japan, launched the project on EFL writing instruction in 1988. The primary research

interest was focussed on the effect of teacher feedback on EFL writing quantity and quality (see Hatori et al. 1990; Kanatani et al. 1993; Tono and Kanatani 1995 for more details). Throughout the data-collection procedures, we collected the free composition data in English from subjects of different academic backgrounds (eighth grade through twelfth grade) and accumulated the data in a machine-readable form.

Table 1 shows the framework of the corpus. The data collection procedure has been largely dependent on the research design of the original writing project. Therefore, the learner profile does not seem to be entirely systematic. For instance, the data for third-year senior high school students was obtained for the first project in 1989 and the number of the subjects was 280. But the following project in 1993 only allowed for 120 subjects for each grade. The data for SH1 was not obtained at the time of the second project in 1993, because the primary focus in our original project was to see if teacher feedback on writing did make a difference as the academic grades increased. This is also the reason why we did not obtain the data for SH1 which would enable us to see the differences between JH groups and SH groups clearly. We will have to fill the gap by collecting the data for SH1 in the near future. We should also note that the data available for the present study was limited to the subcorpora except for SH3. Because of learner profile database management problems, we could not use SH3 data for the analysis of collocation errors.

The size of the whole corpus is about 0.7 million words. As can be seen in Table 1, the size becomes larger for upper-grade groups because more advanced students wrote longer essays. We will have to collect more data for lower-grade samples to create a balance in size among the different academic grade sub-corpora. This corpus is one of the largest learner corpora available in Japan and probably one of the first attempts in the world to collect the interlanguage data from different developmental stages.

3. Collocation errors of English basic verbs

In order to see how the learner corpora contribute to L2 dictionary-making, let us look at the actual data taken from the corpora and discuss its application for lexicographical description. Since it was impossible to examine all the lexical items in the corpus, we chose basic verbs and their collocations for analysis. Table 2 shows the list of the verbs used for the study. As Sinclair says, in order to study the behavior of words in texts, we need to have available quite a large number of occurrences. "About half of the vocabulary of a text — even a very long text — consists of words that have occurred once only in that text." (Sinclair 1991: 18)

In order to get statistically meaningful results, we have to obtain enough observations for each lexical item. In this sense, it was difficult to deal with lexical items whose frequencies were relatively low. Since the size of our

learner corpus was around 0.7 million words, it was almost impossible to have enough occurrences of each of the basic verbs listed in Table 2.

Another alternative was to choose high frequency words such as *the, of, and, to, a, in, that, I, it,* and so forth. They are the forms which occur so frequently that there is no problem to apply statistical procedure to those items. However, as can be seen, most of them are so-called function words and the behavior of these words is rather fixed. We thought that it would be more interesting and of more central importance to include basic verbs in our scope. This does not necessarily mean that the study of those functional words is unimportant. We would like to deal with those items in future research.

4. Corpus analysis procedure

4.1 The basic procedure of text processing

The basic procedure of text processing is shown below:

1) Data Input
2) Preprocessing
   i) SGML removal
   ii) Dehyphenation
   iii) Tokenisation
   iv) Morphological Analysis
3) Tagging (if necessary)
4) Low-level Parsing (if necessary)[3]

(Grefenstette 1995)

In our analysis, we did not have to preprocess the data as such because they were not taken from electronic sources or OCR. The problem, however, is that we transcribed the composition in a Japanese word-processed format, so all the Japanese characters were typed in Japanese. This made it difficult to compare our data with the data from the Bank of English by PC-DOS programs such as LEXA. All the Japanese characters were just unrecognizable on the program.

Tagging and low-level parsing are necessary steps that must be taken in order to investigate the syntactic behaviors of the words in depth, but in this instance we could not use these procedures. The main reason was that the normal tagger or parser did not work correctly on erroneous texts. Therefore, if you are serious about tagging the learner data, you have to do it manually, or first run the automatic tagger and then correct the text manually. This will be one of the biggest obstacles for further research in this area[4]. Very few studies have been done on how to systematically tag erroneous texts (see, however, Meunier forthcoming). We believe that we will have to overcome this problem in order to fully appreciate the benefit of learners' corpora.

We obtained three different types of statistics for each verb lemma. Here we have to clarify the use of the terms. A lemma is what we normally mean by a 'word.' Many words in English have several actual word-forms — so that, for example, the verb *to give* has the forms *give, gives, given, gave, giving,* and *to give.* In this text, the composite set of word-forms is called the lemma. This definition is based upon Sinclair (1991: 173). The three statistics are frequency score, MI-score and T-score. Let us take a closer look at each of these.

4.2 Frequency scores

The simplest way to look at the corpus data is to get a frequency list, i.e. how often each different word-form occurs in the text. There are a couple of ways to arrange the list. Sinclair (1991: 30-31) has described three ways: first, turning the text into a list of the word-forms in the order of their first occurrence, noting the frequency of each; second, sorting it in an alphabetical order; third, sorting in a frequency order. In either case, it is very easy to compare relative occurrences of each word.

Let us look at an example. Table 3 shows a part of the frequency list of the verb *bring* in our learner corpus. This data simply tells us that the most frequently occurring words with the lemma *bring* are *my* (69 times), *out* (62 times), *a* (26 times), *the* (12 times) and so forth. It indicates that the learners use this verb with noun phrases and phrasal verbs such as *bring out.*

4.3 Mutual Information Statistic

The mutual information statistic was first introduced for corpus analysis by Church and Hanks (1990). It basically works as a tool for identifying interesting associations among words in a corpus. Suppose that we saw the sequence "bring a " showing up a number of times in the concordances to BRING and wanted to know if there might be a linguistically interesting pattern. Some sequences in the concordances are interesting (e.g. *bring out*), but others such as *bring a* are not, even though they may be quite frequent. Mutual information can help distinguish the more interesting sequences from the less interesting ones by comparing the joint probability of the sequences with chance. Pairs of words with high mutual information scores are likely to be interesting to a researcher. (For more details, see Church et al. 1991; 1994) Table 4 shows the mutual information statistic for *bring.*

4.4 t-scores

The t-scores compare probabilities that "a third word co-occurs with either of two-words" (Grefenstette 1995: 61). For example, we are interested in which is

more common to say, *powerful tea* or *strong tea*. The t-scores will statistically examine which words are significantly more likely to appear after *strong* than after *powerful* (Church et al. 1991: 125). Table 5 shows the t-scores of the verb *bring*.

4.5 Procedures for collocation data analysis

After choosing the basic verbs, we first obtained the frequency lists of each kind of verb form. Next we picked up collocation errors from the lists. Since we had not tagged all the texts yet, we could not pick up errors according to the parts of speech information. Instead, we identified the errors by looking at the first words which immediately followed the node words (in this case, verb lemmas).

5. The results of verb collocation analysis

Table 6 indicates the relative frequency of the basic verbs selected for our study. The verb selection was made according to the frequency data of an English learner's dictionary. The frequency list indicates that even though we chose 70 different verbs, more than a half of them could not actually be used for our study[5]. For example, the verb *carry* occurred only 15 times in the whole learner corpus data. It is very unlikely that any interesting error pattern would appear in such small samples. If we try to generalize any particular pattern by statistically judging its probability, then we need at least more than 10 expected frequencies in each cell[6]. In our case in Table 6, only a small number of verbs such as *become, bring, come, go, get, have, make, play, see, take, think,* and *want* meet this condition.

Table 7 shows the list of verb collocation errors. The number of the learner errors obtained from the individual composition tasks was rather limited. The main reason for this is that for our free composition tasks, we did not use a multiple-draft design in which the subjects were asked to rewrite the same drafts again and again. Instead, we used different topics for each writing task. Therefore, it was more difficult to collect the data of the same error patterns or corrected forms of the same verbs in different compositions. In spite of the difficulties in data collection, it still indicates some interesting error patterns of the basic verb collocations. We will discuss the results and implications for L2 dictionary making.

6. Integrating the error information into lexicographical description

The learner corpus data shows that the learners fixed error patterns in their use of verb collocations. There are many possible sources of errors such as interlin-

gual errors (overgeneralisation from L1 structures or semantic or lexical structures) or intralingual errors (overapplication of L2 rules, etc.). Whatever the sources, it would be useful for the learners of English to find the information on frequently occurring error patterns. Let us look at some of the common error patterns for EFL learners in Japan and how we could integrate such error information into the dictionary design.

6.1 Errors of verb meanings

The results show that the learners had a tendency to use wrong verbs which were quite similar in meaning. **Become to do*, for instance, is a literal translation of the Japanese phrase "suru youni naru". Learners usually learn the meaning of *become* as "naru" and *come* as "kuru". For Japanese learners of English, the word *become* is more strongly associated with the phrase "suru youni naru" (*come to do*) than the word *come* itself. This kind of error is caused by L1 transfer of verb meanings. The same type of error was observed in the phrase such as **look a dream* (in Japanese, is used the verb *miru* (look; watch) for "have a dream") or **take concert* (which means "have a concert").

In L2 lexicography, therefore, it is very important to provide usage notes on frequently occurring errors such as **become to do* under the entry *come* or *become*. Such learner errors have been ignored in describing a lexical entry, but if it is designed for language learners, the dictionary should contain such information in problematical areas for learners.

6.2 Errors of verb patterns and collocations

The data also shows that learners make quite systematic errors in the use of prepositions or particles after verbs. For instance, many subjects dropped the prepositions in phrases such as "come *to* ...", "come back *to* ...", "go *to* ...", "look *at* ...", "think *about* ..." and so on. In Japanese, no prepositions are needed for these verb expressions, so this might be another case of L1 interference. It is also quite confusing for Japanese learners of English that some of these verbs could be used without the prepositions if the following elements are adverbials (e.g. *Come here. Come back here. Go home.*). Therefore, knowing which prepositions or particles should follow the verbs is also another problematical area for Japanese learners of English.

Another common error is to use wrong verb patterns. For example, **go to shopping* instead of saying *go shopping*, or **want do* for *want to do*. These grammatical patterns are very complicated for Japanese learners and they have to learn the behaviors of each verb one by one. Currently most bilingual dictionaries in Japan and monolingual learners' dictionaries such as LDOCE or COBUILD all provide useful grammatical codes for these verb patterns. The

information on the most difficult verb patterns for certain groups of learners, however, has not been fully investigated and described in a dictionary.

In pedagogical dictionaries, more and more information on these collocation errors of "verb + preposition / particle" or other verb patterns should be systematically provided. Especially, the learners should be warned of not only **possible** errors but also **frequent** errors by collecting more data on learner English. For advanced learners, the collocation information for the verbs or nouns at 5000 to 7000 word levels is very important, but not many dictionaries offer useful information in a systematic way for this level of lexical items.

7. Conclusion

So far we have seen how the learner corpus can contribute to the systematic analysis of learner errors and how those errors should be dealt with in dictionaries. The effect of negative evidence (i.e. the information on 'something is not possible') in dictionaries is to be empirically tested, but it is worth noting that the information on L1-related errors or the most frequently occurring errors can provide the L2 dictionary users with useful guidelines for correct usage.

Some bilingual (English-Japanese) dictionaries in Japan contain this kind of negative information, but there are still many editors and lexicographers who have reservations about providing "incorrect" usage in a dictionary. This question, however, is worth investigating empirically and more attempts should be made to improve the design of pedagogical dictionaries in order to best suit the needs of language learners.

Notes

1. Longman is said to have about 8 million words in its learner's corpus. (P. Scholfield, personal communication)
2. Milton (personal communication).
3. Some researchers classify tagging and parsing as one of the preprocessing stages (for instance, Church et al. 1991). This stage was based upon the lecture given by Gregory Grefenstette in the Seminar on Computational Lexicography at Kossuth Lajos University in Debrecen, Hungary, from Nov. 27 to Dec. 1, 1995.
4. I would like to thank Fanny Meunier for her helpful comments on the problem of error tagging.
5. These frequency scores were based upon the data from JH-2, JH-3, and SH-2 because of the technical problems we had at the time of data analysis in 1995. This is why the frequency scores were rather small in size.
6. Brown (1988: 190). Usually the occurrences of certain lexical patterns are regarded as frequencies of certain categorical variables. Therefore, non-parametric analysis such as chi-square or phi-coefficient is suitable.

References

Asao, K., K. Nozawa, S. Ozeki, M. Sugiura and K. Harada. 1995. Learners' Corpus and Second Language Acquisition Research. Paper presented at the symposium at 34th Annual JACET Convention, Seijo University, Tokyo.

Brown, J.D. 1988. *Understanding Research in Second Language Learning*. Cambridge University Press.

Church, K. and P. Hanks. 1990. Word Association Norms, Mutual Information and Lexicography. *Proceedings of the 28th Annual Meeting of the Association for Computational Linguistics*: 76-83.

Church, K., W. Gale, P. Hanks and D. Hindle. 1991. Using Statistics in Lexical Analysis. U. Zernik (Ed.). *Lexical Acquisition*: 115-164. New Jersey: Lawrence Erlbaum Associates, Inc.

Church, K., W. Gale, P. Hanks, D. Hindle and R. Moon. 1994. Lexical Substitutability. B.T.S. Atkins and A. Zampolli (Eds.). *Computational Approaches to the Lexicon*: 153-177. Oxford: Clarendon Press.

Granger, S. 1994. The Learner Corpus: a Revolution in Applied Linguistics. *English Today* 39, vol.10 (3): 25-29.

Grefenstette, G. 1995. *Computational Lexicography*. Paper presented at the Seminar on Computational Lexicography held at Kossuth Lajos University, Debrecen, Hungary, November 27 - December 1, 1995.

Hatori, H. et al. 1990. *Effectiveness and Limitations of Instructional Intervention by the Teacher. Writing Tasks in EFL*. Tokyo: The Ministry of Education.

Kanatani, K., K. Itoh, T. Noda, Y. Tono and N. Katayama. 1993. *The Role of Teacher Feedback in EFL Writing Instruction*. A research project report, Grant-in-Aid for Scientific Research (B). Tokyo: The Ministry of Education, Science and Culture.

Meunier, F. Forthcoming. Tagging and Parsing Interlanguage. L. Beheydt (Ed.). *Linguistique appliquée dans les années* 90: 21 - 29.

Sinclair, J. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Tono, Y. and K. Kanatani. 1995. EFL Learners' Proficiency and Roles of Feedback: Towards the Most Appropriate Feedback for EFL Writing. *Annual Review of English Language Education in Japan* 6: 1-11. The Federation of English Language Education Societies in Japan.

Table 1: Framework of TGU Learner Corpora Project

Learner Profile:

Academic year		Number of Subjects
1) 8th grade	(JH 2)[1]	120
2) 9th grade	(JH 3)	120
3) 10th grade	(SH 1)[2]	—[3]
4) 11th grade	(SH 2)	120
5) 12th grade	(SH 3)	280

Learner Profile:

All the writing tasks were free compositions on the following topics:

1) Which would you prefer, rice or bread for breakfast?
2) What would you bring out if fire broke out?
3) Write the story of "After Urashima Taro"[4]
4) Tell us about your most horrible dream.
5) Tell us about your school festival.
6) Tell us about your summer holidays.

Corpus size:

1) 8th grade	(JH 2)[5]	96,696 words
2) 9th grade	(JH 3)	110,037 words
3) 10th grade	(SH 1)[6]	———
4) 11th grade	(SH 2)	143,942 words
5) 12th grade	(SH 3)	348,340 words
		699,015 words

1 JH 2 = Junior High School 2nd year

2 SH 1 = Senior High School 1st year

3 This project was originally conducted for comparing the junior high school groups and the senior high school groups in term of the effect of teacher feedback on composition drafts. The data for SH 1 has not been taken because of the technical problems of our original research design.

4 *Urashima Taro* is a traditional Japanese folktale. Urashima saved a turtle and it took him to Sea Paradise and there he received a beautiful gift box. When he was back on shore, he opened the box and became an old man. The students were asked to write what happened to Urashima after that.

5 JH 2 = Junior High School 2nd year

6 SH 1 = Senior High School 1st year

Table 2: The list of the verbs used in the study

ask	knock	set
break	know	settle
bring	lay	shake
call	leave	show
carry	live	sit
change	look	speak
come	make	stand
count	mind	start
cut	move	stop
do	open	strike
draw	pass	take
drive	pay	talk
drop	pick	tell
fall	play	think
feel	pull	throw
find	push	touch
get	put	turn
give	rise	want
go	roll	wash
hang	run	wear
help	say	wish
hold	see	work
keep	send	write

Table 3: Frequency lists of the verb *bring*

-3		-2		-1		0	1		2		3	
i	17	i	178	will	119	bring	my	69	my	25	and	25
so	16	want	9	would	46	bring	out	62	money	14	i	19
happened	4	so	5	to	31	bring	a	26	bankbook	9	but	9
be	4	will	5	i'll	20	bring	the	12	a	7	of	9
and	3	if	4	i	8	bring	some	6	because	6	first.	9
earthquake	3	able	4	should	3	bring	all	6	of	5	my	8
if	2	and	3	won't	3	bring	it	5	bank	5	because	5
fire	2	forget	2	must	3	bring	money.	4	bankbook.	5	if	5
mustn't	2	time	2	can't	2	bring	money	4	cds	4	book	4
time	2	like	2	not	2	bring	to	4	but	4	is	4
a	2	way	1	be	2	bring	nothing	4	it	3	money	4
probably	2	heavy	1	help	2	bring	them	4	out	3	from	3
too	2	well.	1	he	2	bring	much	3	nothing.	3	a	3
secondly	1	earth-quakei	1	i'd	2	bring	something	3	lot	3	card	3
火事	1	things	1	don't	1	bring	anything	3	first	3	it	3
expemsive	1	have	1	also	1	bring	nothing.	3	i	3	first	2
the	1	food	1	may	1	bring	out.	2	money.	3	in	2
thing	1	decided	1	never	1	bring	mathematical	1	books	3	too.	2
get	1	anything	1	is	1	bring	mu	1	guitar.	2	the	2
occured	1	doraemon.	1	been	1	bring	stamp	1	anything	2	new	2
january	1	happened	1	it	1	bring	moneys.	1	food	2	soccer	2
sleep	1	memo.	1	could	1	bring	file.	1	old	2	old	2
300	1	heavey	1	usually	1	bring	things	1	when	2	photo	2
i'd	1	house	1			bring	precious	1	and	2	too	2

Table 4: List of MI-scores of the verb *bring*

-3		-2		-1		0	1		2		3	
wure	15.561	earth-quakei	15.561	will	12.994	bring	mathematical	15.561	cassettes	15.561	teddy	15.561
expemsive	15.561	momet	15.561	would	12.986	bring	bandbooks	14.561	cardboards	15.561	kit	15.561
kichen	15.561	heavey	14.561	i'll	12.635	bring	mu	13.976	per-sonal	15.561	panick	15.561
occured	15.561	ready	12.754	i'd	11.976	bring	coco's	13.239	outdoor	15.561	player's	15.561
ofcurse	15.561	afford	12.239	won't	11.898	bring	stamp	13.239	impoatant	15.561	zack	15.561
mustn't	14.239	able	11.452	should	11.473	bring	everywhere	12.976	moneybank	15.561	instrument	15.561
havea	13.976	forget	11.169	help	10.584	bring	bird	12.754	mental	14.561	elder	14.561
gets	13.561	happened	10.806	must	10.135	bring	out	12.529	walk-man	14.561	file	13.976
secondly	13.561	heavy	10.607	may	9.754	bring	precious	12.239	platteriots	13.976	remind	13.976
probably	13.239	try	10.313	also	9.169	bring	personal	10.86	attack	13.976	uniform	13.976
ufo	13.239	decided	10.239	been	9.118	bring	my	10.809	electric	13.976	photo	13.754
happened	12.806	minutes	10.169	to	8.994	bring	anything	10.736	bento	13.561	pass	13.561
there're	12.561	i	9.942	could	8.754	bring	album	10.703	keyboard	13.561	violin	13.561
january	12.239	want	9.495	never	8.642	bring	nothing	10.678	key	13.561	card	13.561
perhaps	12.101	way	9.169	can't	8.308	bring	others	10.607	lighter	13.561	pet	13.561
mine	11.561	anything	9.151	be	7.98	bring	something	10.203	shrine	13.561	mountainering	13.561
i'd	10.976	need	9.085	not	7.854	bring	some	10.022	note	13.239	palace	13.239
changed	10.806	if	8.712	usually	6.86	bring	all	9.991	compact	13.239	bass	12.754
earthquake	10.754	food	8.572	don't	5.995	bring	them	9.672	photos	13.239	discs	12.561
once	10.432	will	8.421	he	5.911	bring	two	9.186	coins	13.239	6	12.391

Table 5: List of t-scores of the verb *bring*

-3		-2		-1		0	1		2		3	
i	4.079	i	13.328	will	10.907	bring	my	8.302	my	4.992	and	4.986
so	3.992	want	2.996	would	6.781	bring	out	7.873	money	3.738	i	4.317
happened	2	will	2.23	to	5.557	bring	a	5.09	bankbook	2.999	of	2.99
be	1.996	so	2.221	i'll	4.471	bring	the	3.444	a	2.628	but	2.988
earthquake	1.731	able	1.999	i	2.764	bring	some	2.447	because	2.444	my	2.815
and	1.692	if	1.995	won't	1.732	bring	all	2.447	bank	2.235	if	2.232
mustn't	1.414	and	1.692	should	1.731	bring	it	2.221	of	2.223	because	2.23
probably	1.414	forget	1.414	must	1.731	bring	nothing	1.999	cds	1.999	book	1.999
fire	1.413	time	1.409	i'd	1.414	bring	them	1.998	but	1.982	money	1.993
too	1.411	like	1.405	help	1.413	bring	money	1.993	books	1.73	is	1.98
time	1.409	earth-quakei	1	can't	1.41	bring	to	1.97	lot	1.73	card	1.732
if	1.407	momet	1	be	1.409	bring	anything	1.731	first	1.73	from	1.73
a	1.381	heavey	1	not	1.408	bring	something	1.731	out	1.726	it	1.712
occured	1	ready	1	he	1.391	bring	much	1.728	it	1.712	a	1.705
kichen	1	afford	1	may	0.999	bring	mathematical	1	i	1.628	photo	1.414
expemsive	1	happened	0.999	also	0.998	bring	bandbooks	1	lighter	1.414	soccer	1.414
wure	1	heavy	0.999	been	0.998	bring	mu	1	shrine	1.414	bank	1.413
ofcurse	1	try	0.999	could	0.998	bring	coco's	1	cat	1.414	new	1.413
havea	1	decided	0.999	never	0.997	bring	stamp	1	pillow	1.414	first	1.411
secondly	1	minutes	0.999	usually	0.991	bring	everywhere	1	purse	1.414	too	1.411
gets	1	way	0.998	don't	0.984	bring	bird	1	album	1.414	old	1.41
ufo	1	anything	0.998	it	0.966	bring	precious	1	tennis	1.413	or	1.41
there're	1	need	0.998	is	0.96	bring	personal	0.999	anything	1.413	can't	1.41
january	1	food	0.997			bring	album	0.999	what	1.412	when	1.405

Table 6: The frequency list of the verbs used in the study

ask	25	knock	2	set	16
break	75	know	128	settle	0
bring	254	lay	3	shake	9
call	61	leave	63	show	42
carry	15	live	159	sit	15
change	38	look	115	speak	14
come	256	make	275	stand	27
count	2	mind	22	start	33
cut	9	move	33	stop	20
do	272	open	65	strike	0
draw	1	pass	10	take	459
drive	4	pay	19	talk	37
drop	1	pick	0	tell	51
fall	64	play	280	think	508
feel	146	pull	4	throw	12
find	113	push	5	touch	2
get	411	put	27	turn	22
give	120	rise	9	want	558
go	480	roll	2	wash	6
hang	0	run	125	wear	13
help	80	say	154	wish	21
hold	41	see	219	work	60
keep	43	send	10	write	75

Table 7: The list of collocational errors of the basic verbs

Errors	n	Examples (sic)
ask for non-NP	3	I might call Dora and ask for *take me to 22th Century.
ask sb V-ing	1	I will ask my brother *helping me.
become to do	10	If there are a lot of money, I only become *to be able to buy a lot of things.
bring + ø	1	I won't be able to bring *because of panick.
bring to NP	4	I want to bring *to it.
change NP	6	It became bigger and bigger and changed *Doraemon_JP.
be changed NP	2	He was changed *his seikaku_JP.
come (no prep) NP	9	Why don't you come *my school festival?
come back NP	8	So he was glad to come *back his house.
come there	7	Many people came *there and very crowded.
do NP	30	I couldn't do *breath / hayaben_JP / jikken_JP.
fall NP	6	I often a dream that I fell *a hall
feel S	2	I don't feel *don't have breakfast every morning.
feel to do	1	I don't feel *to have had a breakfast
get oldness	1	Taro's getting *oldness
get VP-ing	1	I am getting *sleeping sleepier and sleepier
get use to NP	1	we got *use to it when we were going ...
go (no prep) NP	13	I went *cocert this group with my friend.
go to ADV	3	I go *to somewhere to play
go to V-ing	6	When he went *to fishing,
have V-ing	1	I don't have *bringing things very much
have ADJ	1	I had *interested in this year.
help NP	6	I helped *a lot of work.
help V-ing	1	In the band, I help *making posters.
hold (be held)	7	The festival *held for three days.
keep me alone	1	Please keep *me alone.
keep to do	1	If you were to keep *to take no breakfast,
leave at NP	3	I leave *at the home at 6:50
leave to NP	1	Everyone left to school.
live NP	2	One day I was living *a wonderous room.
look a dream	1	I don't like to look *a dream

Table 7 (*continued*)

look (no prep) NP	22	One day I looked *her face,
make NP *to do*	1	Who made *people to eat three meals?
play (no prep) NP	2	I used OTOSHIDAMA for buy the clothes and play *city.
be run by NP	3	I was being run *by a bad man.
say NP + NP	1	I want to say *my classmate "Thank you".
show to NP	1	I want show *to everyone as many as I could.
sit (no prep) NP	4	I can't sit *the seat in train.
speak (no prep) NP	1	He spoke *many people about his adventure.
stand (no prep) NP	1	I hope that I will stand *the stage.
stop (no prep) NP	1	That trains don't stop *station that I was there.
take (wrong) NP	1	A American rock band come to Japan and take *concert.
take NP (no prep) NP	3	he takes him *Ryugujo again.
talk NP	2	He wouldn't talk *anyone else about his story
tell NP NP	1	I must not tell *others it.
think (no prep) NP	15	I can't think *all things
think VP	8	I thought *buy new ward-processer
turn back + Adj	1	he tried to turn *back young,
turn NP	3	After he turned *old man,
want VP	15	I want *be tall
wish if ...	1	I wish *if there were some.
wish do	2	he wish *return to Ryugujo.
be written NP	2	A paper on my desk is written *"look behind you."

Note: The column n indicates the number of subjects who made these errors.