

Between Verbs and Nouns and Between the Base Form and the Other Forms of Verbs – A Contrastive Study into COLEC and LOCNESS

Xiaotian Guo
Birmingham University

Introduction

One of my previous contrastive studies into a learner corpus COLEC¹ and a native speaker corpus LOCNESS² has shown that these learners overuse some of the verbs. In this article, I am going to apply the theories of Halliday and Biber et al to find out some possible reasons. There are two questions I am going to answer throughout this paper: Do COLEC writers use verbs more than LOCNESS writers where there are noun equivalents? And do COLEC writers use the same amount of base form as LOCNESS writers within the verbs? In the first part, a study of twenty-five verbs and their related nouns is carried out into the tendencies of the corpus writers of the two groups in choosing nouns in their writings. In the second part, a study of fifty verbs and then four verbs in particular is made focusing on the distribution of the frequencies and the behaviour of the base form and the other forms of the verbs.

I. Verbs or Nouns: A Study of Tendency

According to Halliday (1985: 72-75), it is a characteristic of written English that “lexical meaning is largely carried in the nouns”. He attributed two reasons to this phenomenon: one is the structure of the nominal group and the other is the structure of clause. In analyzing the structure of the nominal group, Halliday believes that “there are a lot of things that can only be said in nominal constructions: especially in registers that have to do with the world of science and technology, where things, and the ideas behind them, are multiplying and proliferating all the time”. He compares the structure of verb groups with that of nominal groups and points out that in verbal groups there is only one lexical element: the verb itself. Even though “other lexical material may be expressed in adverbial groups”, “there are very limited in scope”. In the structure of the clause, Halliday also finds out the internal requirement for nominal groups in the grammar of modern English. Based on corpora investigations, Biber et al (1999:65) finds out that the lexical word classes vary greatly both in overall frequency and across registers. In overall frequency nouns are the most frequent word class and across registers nouns are most common in news and academic prose but least in conversation.

As has been found in my previous study of COLEC, learner English is characterized by spoken-like language rather than an academic style which is supposed to be the right one. Supported by this finding and enlightened by Halliday and Biber et al’s theories, it can be predicted that native speakers will tend to use nouns more frequently than learners where there are noun equivalents of verbs.

Now I will choose some verbs and their noun equivalents to find out whether it is true that native

¹ COLEC is a corpus of learner English mainly composed of Chinese university students’ essays in national tests. Its size is 503,799 words.

² LOCNESS is a corpus of native speakers. It is composed of four components, namely, British Essays of A-Level students, British essays of university students, American argumentative essays and American literary-mixed essays. Its size is 324,161 words.

speakers tend to use nouns more frequently than learners where there are noun equivalents of verbs. If this is true, it means the tendency to choose noun equivalents is stronger for native speakers than for learners. These verbs and their noun equivalents are as follows:

accept (acceptance), apply (application), argue (argument), assume (assumption), believe (belief), choose (choice), commit (commitment), communicate (communication), compare (comparison), complete (completion), create (creation), enter (entry), examine (examination), express (expression), include (inclusion), indicate (indication), introduce (introduction), involve (involvement), manage (management), occur (occurrence), produce (production), realise (realization), realize (realization), refuse (refusal) and survive (survival).

Table 1 shows the frequencies of these verbs and their equivalent nouns. Each verb in the table is referred as lemma including all the forms of the verb: the base form, the third singular form, the “-ing” form, the past form and the past participle. Each noun frequency includes both the singular form and the plural form.

Table 1: The frequencies of some verbs and frequencies of their equivalent nouns

	COLEC	COLEC	LOCNESS	LOCNESS
	VERB	NOUN	VERB	NOUN
ACCEPT	41	0	182	33
APPLY	65	2	60	11
ARGUE	1	3	167	339
ASSUME	3	0	40	13
BELIEVE	295	14	373	125
CHOOSE	121	31	140	129
COMMIT	8	0	90	12
COMMUNICATE	24	30	24	27
COMPARE	52	2	49	15
COMPLETE	35	0	42	5
CREATE	18	1	182	38
ENTER	84	1	56	10
EXAMINE	17	60	28	5
EXPRESS	25	9	56	12
INCLUDE	67	0	111	2
INDICATE	21	0	10	6
INTRODUCE	12	2	61	44
INVOLVE	12	0	159	9
MANAGE	29	8	27	12
OCCURE	25	0	96	5
PRODUCE	239	65	89	38
REALISE	9	0	98	16
REALIZE	196	3	122	14

REFUSE	27	0	64	13
SURVIVE	34	4	47	16
Total	1460	235	2373	949
N Total ³	28.98	4.66	73.2	29.28

If we look at the total use of these fifty verbs and their equivalent nouns, there are 1460 cases of verb use and 235 cases of noun use in COLEC and there are 2373 cases of verb use and 949 cases of noun use in LOCNESS. If we normalize these figures by the total frequency of the two corpus respectively, we get the following comparison: in every 10,000 words, there are 28.98 cases of verb use and 4.66 cases of noun use in COLEC and there are 73.2 cases of verb use and 29.28 cases of noun use in LOCNESS. If we compare the noun use with the verb use across the two corpora based on normalized total frequency, we will get the following ratio: NOUN:VERB in COLEC is 0.16 and NOUN:VERB in LOCNESS is 0.4. This means LOCNESS writers use two and half times more than COLEC writers in noun use.

If we look at the individual frequency of these verbs and their equivalent nouns, we may find that most nouns are less in number than verbs in both corpora. But there are also a few exceptions. For example, *examination* appears much more frequently than that in LOCNESS because COLEC writers are using a lot the sense of test for *examination* rather than the sense of investigation as can be found in LOCNESS: *an exhaustive examination of the broadcast networks' programming*. Since tests are overwhelmingly a major concern of university students, the overuse of this sense of *examination* is understandable. Another point I have noticed is the frequent occurrences of *production*, which is the largest use of nouns within these nouns in COLEC. This is caused by the topic about the production of fake commodities in a majority of the essays in COLEC. If we may ignore these two cases, however, we will find a more striking underuse of nouns in COLEC.

Table 2: The total frequencies and the normalized total frequencies of the selected verbs and those of their related nouns in COLEC and LOCNESS

	COLEC	COLEC	LOCNESS	LOCNESS
	VERB	NOUN	VERB	NOUN
Total	1204	110	2256	906
N Total	23.9	2.18	67	27.95

If we compare the ratios between the cases of noun use and the cases of verb use across the two corpora, we may see the difference of COLEC writers underuse of nouns is getting extremely larger: NOUN:VERB in COELC is 0.09 and NOUN:VERB in LOCNESS is 0.4. LOCNESS writers are using nouns 4.4 times more than COLEC writers. To put it in another way, the tendency to choose nouns is even milder in COLEC if we may discard the most often used ones due to the influence of topics. After it is certain that verbs are more often chosen by COLEC writers in language use, it would help us to understand better of learner English if we can know how verbs are used between the different forms.

³ "N Total" stands for normalized total. It means the comparable total normalized by the total frequency of the corpus.

II. The Favoured Form of Verbs: the Base Form

As has been found in my previous studies, there is a large discrepancy between the base form and other forms of a verb in COLEC than in LOCNESS. This part will examine the detailed differences of them from the use of the different forms of 50 verbs and then four verbs in particular to see how properly the base form is used in COLEC. These verbs are chosen first and then selected and then re-chosen to make sure that they are verbs only and any individual form of any individual verb does not overlap in spelling with its noun or other lexical form. It has been a difficult process to choose the fifty verbs. For example, KNOW was chosen first but was found to be overwhelmingly used by learners. Thus, it was deleted from the list in the end not to let it deform the whole picture of frequencies and affect the soundness of judgment. Table 3 shows the distribution of these 50 verbs in their base form (V), present participle form (V-ing), third singular form (V-s), past form (V-ed(1)) and past participle form (V-ed(2)). The corpus of COLEC is represented by C and LOCNESS by L.

Table 3: The distribution of all the forms of 50 verbs in COLEC and LOCNESS in alphabetical order

V	V		V-ing		V-s		V-ed(1)		V-ed(2)	
	C	L	C	L	C	L	C	L	C	L
ACCEPT	28	92	5	21	0	14	8	55		
AFFECT	32	29	0	14	7	23	13	37		
AFFORD	16	38	0	0	1	0	1	2		
ALLOW	5	83	1	43	0	47	5	100		
APPLY	51	27	1	6	4	11	9	16		
ARGUE	1	81	0	13	0	29	0	44		
ASSUME	3	21	0	7	0	3	0	9		
BEGIN	108	52	99	64	8	39	52	54	2	22
BELIEVE	269	226	1	13	6	76	19	58		
BRING	290	80	4	18	34	35	37	81		
CARRY	39	48	3	21	2	10	17	40		
CEASE	1	15	0	3	0	0	1	1		
CHOOSE	99	57	5	20	1	23	13	27	3	13
COMMIT	7	36	0	20	0	10	1	24		
COMMUNICATE	18	18	4	4	0	1	2	1		
COMPARE	11	12	11	1	1	4	29	32		
COMPLETE	32	34	0	1	0	0	3	7		
CONSIDER	88	60	7	26	2	9	26	81		
CREATE	11	70	1	28	0	19	6	65		
EAT	127	44	29	50	3	1	13	1	4	8
ENTER	60	31	8	7	1	8	15	10		
EXAMINE	12	13	2	5	1	2	2	8		
EXPRESS	22	24	1	7	1	6	1	19		
GIVE	297	164	10	62	31	52	23	41	42	147
GO	684	212	169	148	63	92	115	35	22	37

HAPPEN	69	66	22	25	13	35	76	30		
HURT	60	23	0	7	3	4				
INCLUDE	16	38	42	39	7	17	2	17		
INDICATE	5	4	1	3	12	2	3	1		
INTRODUCE	5	13	2	8	0	2	5	38		
INVOLVE	5	17	1	16	0	24	6	102		
LEAD	138	130	10	30	33	49	25	74		
MANAGE	22	4	2	0	0	10	5	13		
OCCUR	13	47	2	9	5	24	5	16		
PRODUCE	145	48	46	12	12	10	36	19		
REALISE	9	33	0	7	0	38	0	20		
REALIZE	137	72	8	15	1	21	50	14		
REFUSE	19	12	1	10	0	23	7	19		
RUN	126	84	44	52	5	12	30	7		
SAVE	199	40	32	29	5	8	11	10		
SEEM	18	130	0	2	48	140	22	24		
SEND	20	13	16	3	0	3	19	20		
SPEAK	161	35	134	25	3	4	3	4	33	4
SPREAD	6	19	3	4	2	1				
SURVIVE	34	31	0	6	0	3	0	7		
TAKE	909	293	99	115	51	77	63	59	117	136
TELL	151	41	9	11	41	41	77	57		
THROW	17	13	5	7	1	4	1	1	6	22
WANT	1154	217	5	17	94	106	99	90		
WRITE	228	19	105	25	0	30	7	23	17	33
Total	5977	3009	950	1079	502	1202	963	1513	246	422

If we add the frequencies of all the forms of the 50 verbs in COLEC (5977+950+502+963+246), the figure amounts to 8638 and if we add those in LOCNESS (3009+1079+1202+1513+422), the figure is 7225. To get an overall percentage of the use of the fifty verbs (including all the forms of each lemma) in the two corpora, we may divide the two figures above by the total frequencies of the two corpora respectively. For COLEC it is $8638/503799=0.017$ and for LOCNESS it is $7225/324161=0.022$. There is not a dramatic difference between the two groups of corpus writers in the total percentages of verbal uses because most of the verbs are commonly used in both of the corpora. However, if we compare the percentage of the use of the base form and that of the other forms of the verbs in the corpora, the picture will look different (see Table 4). The base form takes up 69.4 percent of the total use of the lemma in COLEC and 41.6 percent in LOCNESS. Contrary to the overuse of the base form of verbs, COLEC writers are underusing all the other forms of verbs.

Table 4. Percentages of all the forms of the 50 verbs in the two corpora

Form	COLEC	LOCNESS
V	69.4	41.6

V-ing	10.9	14.9
V-s	5.8	16.6
V-ed (1)	11.1	20.9
V-ed (2)	2.8	5.8
Total	100	99.8

There may be several reasons for the overuse of the base form of verbs. One possible reason is that COLEC writers are misusing the base form for other forms that native speakers use. This assumption is based upon one of the features of the native language of the learner corpus writers, namely, Chinese is a non-inflected language. COLEC writers are not properly aware of the requirement of the English language in the change of word forms in different contextual situations. In the following part, this assumption will be tested. To do this, four verbs are chosen to see how properly the base forms are used in COLEC: APPLY, HAPPEN, CHOOSE, and BEGIN.

An investigation into all the concordance lines of the base form of these verbs shows several types of misuses of the base form. These misuses can be mainly classified into the following categories:

1. the misuse of the base form as the third singular form. Some examples are:

*"Practice Makes Perfect " also apply to other subjects,
The phenomenon often happen in people's everyday life,
That often happen to everyone student.
Whatever happen, wherever we are, we must keep our mind clear.
So, when he choose a job he must ...*

2. the misuse of the base form as the "-ing" form. Some examples are:

*Because failure is the thing which is always happen
everything has a process from happen to developing
We can know what happen in our city from what we talk about.
"OH, my dear, what's happen? Asked the father.
A good begin is half done.*

3. the misuse of the base form as the past participle form. Some examples are:

*we shall know what has happen outside
We can learn what have happen in society.
Since we have choose the professinl field,
Once I have choose a kind of job.
Well begin is half done.*

4. the misuse of the verb as a noun. Some examples are:

*I have written the apply form to join the Party.
Therefore, choose a good job is very important.*

Now we have no choose, we must ...

...from begin to end ...

There are 6 misuses in APPLY, 16 in HAPPEN, 6 in CHOOSE and 12 in BEGIN. All these misuses take up 11.7 percent of all the lemma of the four verbs. Due to the misuse of the base form for other forms of verbs, the number of the base form use has been improperly presented. Till now, the assumption has been proved to be valid after the investigation into the use of the base form of HAPPEN, CHOOSE and BEGIN, which means COLEC writers do have problems in choosing the proper form of the verb when it is needed. By over-applying the wrong form of verbs, the frequency of the base form has accordingly increased. But the over-applications of the wrong form of verbs vary from verb to verb.

III. Some miscellaneous findings

Apart from the misuse of the verbal forms, there are quite a few cases of misuse of the verb where another verb or even word class should be used. Examples are:

<i>The factory can produce all kinds of products which <u>apply</u> good service for the people.</i>	(provide)
<i>the hospital <u>apply</u> the people with modern service.</i>	(provide)
<i>factories require the adequate water to <u>apply</u> the contemporary production</i>	(meet the need of)
<i>The numerous wells can also <u>apply</u> lots of fresh water</i>	(supply)
<i>We should create many conditions to cause the fire to <u>apply to</u> its advantage</i>	(exert)
<i>more and more people want to live comfortable and <u>happen</u></i>	(happy)

In most of the examples above, the verb APPLY is used to mean SUPPLY. The reason for this kind of misuse is simply due to the similarity in the spelling. Even though the misuse of a verb as another verb does not make any difference for the analysis of the reason of the overuse of the base form of verbs, it is a kind of by-product of this study. It shows a sort of progress of language acquisition and the necessity to distinguish one word from another when they share some similarity in form.

Another striking point of the examination is that a couple of words have been found to be underused dramatically by COLEC writers. For example, the lemma ARGUE has appeared 167 times in LOCNESS whereas it has appeared only once in COLEC. Another example is the lemma INVOLVE which has been used 159 times by LOCNESS writers is used only 12 times by COLEC writers. These two verbs rank 14 and 15 respectively in LOCNESS which are near the top of the list but 50 and 44 which are at the bottom of the list in COLEC. Furthermore, if we refer to Table 1 to see whether COLEC writers are using the related nouns of the two verbs, they still remain underused. Therefore, I would speculate that these two verbs are not adequately acquired by this group of learners compared with other verbs in the list.

Another interesting finding is that COLEC writers tend to choose the British spelling for the word REALIZE instead of the American spelling REALISE (REALIZE 196 in total and REALISE 9) in total. This must be the influence of the text books used by COLEC writers and their teachers.

Conclusion

This chapter has revealed learners tendency to choose verbs where native speakers could prefer to choose nouns for the writing purpose. The ignorance is explicit of the difference of language styles is obvious in the choice of nouns and verbs by COLEC writers. Another important finding of learner English style is that COLEC writers overuse the base form of verbs as a whole but underuse the V-ing form, the third singular form, the past form and the past participle form of verbs. Among the overuses of the base form, a large frequency of misuses of the base form has been found where another form should have been used. COELC writers sometimes choose a wrong word to mean something that should be expressed by another word even though this is not a dominant feature of the misuse. Attention should be to these features of learner English in the field of English learning and teaching.

Reference

1. Biber. D., Johansson S., Leech G, Conrad S. and Finegan E. 1999 *Longman Grammar of Spoken and Written English*. London, Longman.
2. Halliday, M.A.K. 1985 *Spoken and Written Language*. Oxford, Oxford University Press.