Korean Journal of English Language and Linguistics
Vol. 18, No. 4, Winter 2018, 423-441
10.15738/kjell.18.4.201812.423

What English Verb Errors do Korean University Students Make? A Corpus Examination of Learner-Produced Verb Phrases

Kevin Parent
(Korea Maritime and Ocean University)

Parent, K. 2018. What English verb errors do Korean university students make? A corpus examination of learner-produced verb phrases. *Korean Journal of English Language and Linguistics* 18-4, 423-441. This article examines a corpus of writings by Korean university students with an eye towards the verb phrase patterns they use. Provided are a list of verb patterns which they do not use; ones they use but are not, or very rarely, found in writings produced by native speakers; ones they overuse; and ones they underuse. The results show particular problems with the passive voice and the perfect aspect, problems that are magnified when these two are combined. In terms of overuse, it reveals that the simpler verb patterns tend to be used much more frequently than would otherwise be expected. One-word verb phrases (*do/did, has, is* and lexical verbs like *help*), basic negative forms (*do/did not have*), basic modal phrases (*can help, could not help*) and others are overused. Underused verbs include, among others, various realisations of passive constructions and phrasal verbs. Pedagogical implications based on these results are discussed.

Keywords: verb phrase, passivity, ergativity, perfective, phrasal verbs, error analysis, corpus linguistics, learner English

1. Introduction

Underuse and overuse are curious phenomena. To some extent, they indicate a lack of mastery of rules. This is easily the case when learners underuse *a/an* or *the*. But there's a marked difference in over- or underusing articles than in doing so with verb patterns. A noun phrase can only be marked by *a*, *the* or the 0-article, but the structure of verb phrases is open; they have an indefinite, or at least very large, number of permutations. There are many patterns to learn: *stop* + infinitive phrase, *stop* + gerund, *make* (noun) *cry* and many others. Verb phrases (VPs) can be extended with conjunctions or infinitive phrases. The overuse of one form may be the

result of underusing one or more other patterns. Given the sheer number of VP realisations, learners have to start constructing VPs before having encountered all the patterns, thus resulting in some being overused, and their use of such patterns may not necessarily be replaced as new forms are learnt. Rare indeed is the learner who can accurately proclaim to have learnt every possible verb phrase pattern; rather, the mental system may feel (and probably *be*) perpetually incomplete and developing. Certain forms may be encountered without being fully acquired and therefore less available in free production and possibly 'avoided'—or at least underproduced. This can be either because the form has not been completely mastered or because the learner is unsure exactly *when* to use it, if not both. So certain verb patterns are going to be overproduced, some will be underproduced, some may be missing (not used, or just never learnt) and others rendered incorrectly (formal errors). This is what the current article examines.

We examine and compare the verb phrases of both learners and native speakers, guided by the following research questions:

1. Are there verb patterns which the learners do not use, and, if so, what are they?
2. Are there verb patterns which the learners use but which native speakers use rarely or not at all, and, if so, which ones?
3. Are there verb patterns which the learners overuse compare to native speakers, and which ones if so?
4. Which patterns, if any, do learners underuse compare to native speakers?

2. Background

2.1. Underproduction and Avoidance

That the English verb phrase is complex and difficult for L2 learners to master will be taken here both as absurdly obvious and as a starting point for more meaningful inquiry. Voice (active or passive), aspect (perfect or progressive), tense (past or present), mood (indicative, interrogative, subjunctive, etc.) and modality (*can, will*, etc.) along with the issues of argument structure and numbers agreement are trappings that learners from any first-language background face.

Underuse is sometimes taken as synonymous with avoidance. This dates back to

Schachter (1974) who noted differences in the production of restricted relative clauses by writers of Arabic, Chinese, Japanese and Persian background and the observation that the Arabic and Persian speakers produced about as many such clauses as do native speakers of English, but that the Chinese and Japanese produced notably fewer. On the grounds that Arabic and Persian, like English, constructed relative clauses to the right of the noun while Chinese and Japanese pre-modify their nouns, she developed the notion of avoidance, claiming that when learners find certain structures difficult, they may employ paraphrase to get the same idea across. (For the issue of avoidance by verbal phrases by Korean learners specifically, see Bjers and Massicotte, 2015).

Li (1996), however, both refutes and refines this in an article called 'Underproduction does not necessarily mean avoidance.' Unlike Schachter, Li distinguishes between 'conscious avoidance' and 'subconscious underproduction.' In one test, 11 adult learners answered questions in English whose answers required them to use restrictive clauses and to translate sentences from Chinese using this construction. This was followed by an interview. In 37 instances in which the participants did not use a relative clause, most said they did not consider using this construction. No learner said they consciously avoided doing so. The participants were given a second test within two days. This time they were instructed that each sentence they produce should have a relative clause; however, three sentences were included that do not have an equivalent Chinese construction. Except for those three instances, all learners were able to provide a relative clause, strongly suggesting, as Li's title spells out, underproduction does not necessarily equate avoidance.

2.2. Learner Errors in Passive Verb Phrases

Regarding specific verb errors, scholars have noted the passivisation of unaccusative verbs by learners of various first language backgrounds. Perlmutter (1978) advances the Unaccusative Hypothesis, and the phenomenon has been noted as well by Ju (2000) and Cowen, Choi and Kim (2003). The latter investigated a corpus of essays produced by Korean learners of English. They found a higher error rate for the subset of unaccusative verbs that can also be transitive such as *change* ('the situation changed' and 'to change a tyre'), but also noted that errors occurred with both types. They also noted the attempted passivisation (43% of the instances, according to my combination of their data) of the stative verb *exist,* suggesting that passivisation is

being overgeneralised to other verbs. Ju (2000), examining advanced Chinese learners of English, finds an interplay of cognitive factors while developing their L2 proficiency, citing, for one, the saliency of conceptualisable agents. The sentence *An accident happened* does not have an agent in a linguistic-based syntactic/semantic analysis, but nonetheless does have an agent at a conceptual level, leading to its possible passivisation by learners. Zobl writes that the non-standard forms created by learners arise from 'the mapping problem posed by ergative verbs' (1989: 216).

2.3. Learner Errors in Phrasal Verbs

The use (and underuse) of phrasal verbs by learners of English has likewise received considerable attention in research studies. Two recent article provide overviews on the subject. Bronshteyn and Gustafson (2015) is the briefer of these but examines the issues of avoidance, instructor knowledge of phrasal verbs, and innovation in the teaching of these forms. Jahedi and Mukundan (2015) provide an extensive overview of the subject, including avoidance (which, they say, make up the majority of the research) and language teaching materials. Dagut and Laufer (1985) is a well-known study on the avoidance aspect. Their participants, Hebrew-speaking university students of various majors, showed a strong preference for single-word verbs over their phrasal counterparts, attributing this to contrasts in the L1 as Hebrew lacks an equivalent structure.

Hulstijn and Marchena (1989) tested whether avoidance was entirely a grammatical concern, or whether semantics had a role to play as phrasal verbs tend to be more specific and their corresponding one-word verbs more general; hence, an unfamiliarity with the more specific meaning of the phrasal version may also lead learners to avoid their use. They administered three tests to Dutch learners of English, based largely on Dagut and Laufer's methodology. The first was multiple-choice test consisting of a one-word verb, a phrasal verb equivalent, and two distractor verbs. Two answers were therefore correct, but one was preferred by native speakers who had also taken the test. The second involved memorisation, in which sentences were written with the phrasal verbs; subjects were asked to memorise the sentences, and, an hour later, fill out an assessment instrument in which the verbs were replaced by blanks. They were asked to state whether they remembered the word or not and to supply it if possible, and if not, supply one with the same meaning. In the third test, the verb of each sentence was omitted but the Dutch equivalent was provided which the participants

were asked to translate into English. Their findings suggest that avoidance of phrasal verbs can occur, first, when the form is perceived as markedly different from the L1; second, when the form is too similar to their L1; and third, as their hypothesis states, when an L2 form is believed to be more specific than another L2 form. In other words, learners may indeed avoid phrasal verbs on semantic grounds.

Koo (2015) also found a semantic effect, one that varied between language groups. Surprisingly, German learners of English used fewer phrasal verbs than did Korean learners (53% avoidance rate versus 26%). The Korean learners, however, primarily avoided figurative phrasal verbs while the German learners tended more to avoid literal ones.

2.4. Corpus Studies of Learner Verb Phrase Errors

One study with similarities to the current undertaking is Can (2017) who did a large-scale corpus-based study on verb errors in the *Cambridge Learner Corpus*, using the texts produced by Turkish learners. This data is already tagged for errors and contains demographic information about each writer including level according to the Common European Framework of Reference for Languages. Can finds that the most common verb-related error is in selection of the proper tense, while other common errors include the wrong form of a verb, missing verbs, errors in agreement, unneeded verbs, inflection, and derivational errors.

To date, however, no one has attempted a corpus-based exploration of the verb phrase in writings by advanced learners. Examining all the verbs, not just ones with errors, in a corpus of essays by Korean university students, this study aims to fill that gap.

3. Methodology

3.1. Part-of-Speech Tagging

The current study takes a different approach than that conducted by Can (2017). Rather than examining the verbs themselves in context, we examine the patterns of verb phrases according to their C5 tags. A modal verb followed by a lexical verb (as

in 'can hope') would have tag sequence VM0-VVI, where 'VM0' denotes a modal verb and 'VVI' denotes a lexical verb in the bare infinitive form. The negative of this ('can not hope') would be represented as VM0-XX0-VVI. (Variants such as *can't* and *cannot* were collapsed and tallied together with their uncontracted form.) The properly formed passive version ('can be hoped') would have the tag sequence VM0-VBI-VVN, denoting a modal followed by the infinitive form of *be* and the past participle from a lexical verb, etc. Thus our data is the sequence of tags without the actual verbs used, although the original verb phrases will be examined when it is necessary to do so. Moreover, while Can (2017) examined VPs tagged as errors, this study examines all VPs, erroneous or not. It focuses on larger questions and will not necessarily reveal errors in tense or agreement.

3.2. Corpora

Two corpora were used in this study. To examine verb phrases produced by native speakers of English, the *British National Corpus* (*BNC*) was employed, albeit in a modified form, described below. For the learner corpus (LC), a corpus of essays produced by learners from a university in Busan, South Korea was used. Written over several years, the essays that comprise this corpus were written in an annual essay contest hosted by the university in which students voluntarily participate. They are given a prompt with no advance knowledge and asked to write an essay on that topic. Topics, different each year, include, for example, how they define success, or what advice they would give to travellers coming to Korea for the first time, etc. This means that all writers of a given year are writing on the same topic. The essays are written under a time limit. Early finishers may re-read and edit their work if desired, but the works are not edited by native speakers or anyone else; each essay is entirely the work of one author.

The LC was tagged for part-of-speech using the C5 tag set. Rather than using the Stanford tagging software, which frequently didn't tag errors in the desired way, this was done in *R,* by copying tags from corresponding words in the *BNC*, modulated by the presence of other verbs (so *included* would be tagged as a past tense verb if by itself but as a past participle if following a form of *have*, etc.) As the tagged LC has been used in previous research, tagging errors had been manually corrected.

The LC was prepared by limiting it to declarative sentences that contained at least one word tagged as a verb. This restricted the data to canonical sentences and mostly

removed titles and brief pseudo-sentences like 'No!' A script was written in *R* to run through the entire *BNC* and mark all sentence for whether or not they fit these criteria.

Another script was written to delimit and collect verb phrases in the two corpora. This was a very complicated script because, while it is frequently sufficient to get any consecutive words tagged as a verb, which would collect examples like *am, may go* and *will have been waiting*, it would not retrieve other desired instances like *beam me up* (because of the *me*). It's also of questionable value to have instances of *come and go* tallied separately from *look, see* and *do*—that is, two conjuncted verbs versus three or more, though the ability to conjoin verbs is something we want to retain. Many special cases had to be dealt with. (In the case of VPs with conjunctions, these were collapsed into a verb-conjunction-verb pattern, regardless of how many verbs were joined serially.) Still, the inclusion of conjunctions opens a Pandora's box since they occur in passive patterns like *were acquitted and released* and *will arrange to come out and see* and many others. Adverbs were removed, so adverb placement is beyond the scope of the current study. Since adverb particles (as in *goes on*) are included, and conjunctions are as well, VPs like *goes on and on* have to be as well, which results in a long list. Further, it is not necessarily correct that all consecutive words tagged as a verb constitute a unified verb phrase, as seen in *The best thing to do would have been to give up*. In this sentence, only the first three words would not be tagged as verbs, yet the remainder is not a single VP, which for our purposes begins with the modal *would.* Thus rules had to be written for how VPs can begin so that the *to do* would not be counted. Noun phrases that interrupt the VP are also problematic. To be sure, we want instances like 'made me study,' with the middle slot included, but phrases like 'want someone to love' are sufficiently reduced to 'want' as the 'to love' is part of the noun phrases headed by 'someone' rather than part of the VP. The entire algorithm is, then, long and complicated, and manual correction of some data was still performed.

A comparison of the learners corpus to the full written component of the *BNC* would not be a profitable exercise since, as the *BNC* is immensely larger than the LC, there are many more chances for rare VP patterns to appear, patterns that, at least in theory, could be seen in the LC if it were large enough. To work around this, sentences from the *BNC* were randomly selected, using the criterion above, to match sentence lengths (by number of words) in the LC. That is, the first sentence in the LC is 27 words, so the first sentence in the sampled *BNC* corpus is a 27-word

sentence randomly selected from the written component of the full *BNC.* Thus the sampled corpus matches not only the word count, but the sentence count and the words-per-sentence tally and, like the LC, were limited to declarative sentences containing a verb. While this puts the two corpora on equal footing, random sampling could still cause rare VPs to be culled, or even not-so-rare ones to be omitted, so two further *BNC* subcorpora were made as well. Sampling was without replacement, so no sentence appears in more than one subcorpora. While a pattern may not occur in one subcorpora, it is very unlikely to be missing in all three. All corpora are 8,705 sentences in length, 128,017 words total, with an average of 14.7 words per sentence (SD: 7.59). These range from one word to 72, with only one instance of each extreme.

3.3. Analysis

The VPs were extracted from the LC and three subcorpora, and their tallies were put into a spreadsheet. There were a total of 1,844 patterns though many of these were rare, and 186 were not found in the *BNC* subcorpora. Also included were columns for subcorpora mean, length (in number of words), examples from the LC and one *BNC* subcorpora (when there was one). Examples were originally randomly chosen but some have been manually replaced with clearer instances. Additional columns included flags indicating whether or not the VP occurs in all three subcorpora, whether it was deemed significant by an application of Fisher's exact (significance is here defined as having a p-value lower than 0.001 in all three subcorpora). Further flags indicate whether that VP is modal, phrasal, progressive, perfective or catenative. A final column was added to count the interactions so that a VP that was modal, passive, phrasal and catenative (as in 'may be said to sum up') would be coded with a 4. The term *interaction,* then, is not used here in its statistical sense. A pattern is said to have an interaction only when this figure is 2 or higher. Illustration 1 provides an example of the data.

This level of alpha was chosen for the Fisher's Exact text primarily because of the paper by Benjamin and his 19 co-authors (Benjamin et al. 2017) which argues for stricter criteria in p-values in the humanities. While they recommend $p<0.005$, we have opted for an even lower alpha level of $p<0.001$. The other reason this level has been chosen is to limit the data to only the most important cases. Less discriminating levels would yield too many patterns for one article to discuss concisely. Thus all

Pattern	Length	Lrn.n	C1.n	C2.n	C3.n	C.mean	Learner.example	BNC.example	All.3.subcorpora	Significant	L.use	Modal	Phrasal	Progressive	Perfect	Passive	Catenative	Interactions
VVD	1	1952	1713	1733	1673	1706.33	concerned	came	+	-	over	-	-	-	-	-	-	0
VBZ	1	1852	1357	1317	1352	1342.00	is	is	+	+	over	-	-	-	-	-	-	0
VVB	1	2689	1137	1200	1221	1186.00	live	walk	+	+	over	-	-	-	-	-	-	0
VBD	1	1065	1120	1148	1127	1131.67	was	was	+	+	under	-	-	-	-	-	-	0
VVZ	1	550	873	870	807	850.00	makes	requires	+	+	under	-	-	-	-	-	-	0
VM0-VVI	2	1158	608	653	691	650.67	can distinguish	may present	+	+	over	+	-	-	-	-	-	1
VBB	1	749	587	583	553	574.33	am	are	+	-	over	-	-	-	-	-	-	0
VBD-VVN	2	123	363	379	387	376.33	was attracted	was designed	+	+	under	-	-	-	-	+	-	1
VM0-VBI	2	312	242	252	253	249.00	would be	would be	+	-	over	+	-	-	-	-	-	1
VM0-VBI-VVN	3	56	199	220	220	213.00	might be spent	must be planned	+	+	under	+	-	-	-	+	-	2
VBZ-VVN	2	96	228	204	192	208.00	is known	is taken	+	+	under	-	-	-	-	+	-	1
VHD-VVN	2	68	182	166	186	178.00	had studied	had come	+	+	under	-	-	-	+	+	-	2
VHB-VVN	2	123	145	144	144	144.33	have stayed	have gotten	+	-	under	-	-	-	+	+	-	2
VHZ-VVN	2	44	152	144	133	143.00	has increased	has proven	+	+	under	-	-	-	+	+	-	2
VBB-VVN	2	83	123	145	123	130.33	is disappointed	is based	+	+	under	-	-	-	-	+	-	1
VVD-AVP	2	74	131	129	126	128.67	found out	galloped back	+	+	under	-	+	-	-	-	-	1
VBD-VVG	2	97	133	99	110	114.00	was strolling	was getting	+	-	under	-	-	+	-	-	-	1
VM0-XX0-VVI	3	197	117	98	102	105.67	could not explain	can not say	+	+	over	+	-	-	-	-	-	1
VHB	1	320	71	100	90	87.00	have	have	+	+	over	-	-	-	-	-	-	0

Figure 1. Example of Data

results can be considered highly significant.

Note that the C5 tag set used by the *BNC* may be a bit fastidious at times. It distinguishes, for example between 'am going to make' and 'are going to make,' a difference that says more about the subject that the VP. Despite the increase to the number of possible VPs, no attempt was made to combine these. An explanation of all the various tags is not included here for reasons of space but is easily googled. 'NNN' is my own shorthand to denote a noun phrase, whether a pronoun or a longer expression like 'the tall man on the left in the blue fedora.' This is necessary for instances in which a noun phrase is part of the VP such as in 'make someone feel' or 'gave it up.'

4. Results and Discussion

4.1. Verb Phrase Patterns Not Used by the Learners

The first question regarded which VPs the learners do not use. There are numerous instances of VP patterns not used by learners but which also occur infrequently in the *BNC* subcorpora. We have no interest in these and will restrict our attention to those zeroes deemed significant by Fisher's Exact. These are either not known, not needed, or avoided (consciously or not) by the learners. The data set for this, then, includes the rows with: (1) a learner frequency of 0; (2) VPs that occur in all three

subcorpora; and (3) p−values of less than 0.001.

At this selective level of significance, we find only three patterns which the learners did not use but whose absences are statistically conspicuous. All three are passive with interactions. One is a three−way interaction, being modal and catenative in addition to passive; another is passive and catenative, and the final one is passive and progressive. The first, exemplified by 'may be expected to allow,' occurs an average of 18 times in each of the three subcorpora while not at all by the learners. The second pattern can be seen in 'is believed to play,' found around 14 times in the subcorpora, and 'is being performed,' the third pattern, is used an average of nine times. These can be seen in Table 1.

Table 1. Verb Pattern Not Used by Learners at p<0.001

Pattern	Length	C1.n	C2.n	C3.n	C.mean	BNC.example	mod	phras	prog	perf	pass	Catena−tive	Interac−tions
VM0−VBI−VVN−TO0−VVI	5	13	20	21	18.00	may be expected to allow	+	−	−	−	+	+	3
VBZ−VVN−TO0−VVI	4	13	15	13	13.67	is believed to play	−	−	−	−	+	+	2
VBZ−VBG−VVN	3	7	8	12	9.00	is being performed	−	−	+	−	+	−	2

4.2. Verb Phrase Patterns Only Used by the Learners

The second question seeks to find which forms the learners use that are rarely, or never, used by native speakers. Here we must use the original, unfiltered, data set since limiting the data to VPs that occur once in all three subcorpora would eliminate any illicit patterns the learners employed. There are, alas, around 140 VP patterns used by the learners but not found in the *BNC* subcorpora, and 86 found once or twice. The low level of alpha we have chosen allows us to filter most of these out. The data we examine here are those rows with (1) a learner frequency of at least one; (2) a mean of subcorpora frequency of ten or less; and (3) a tag in the Fisher's Exact significance column.

We find six such patterns. All are grammatical and found within the subcorpora though not necessarily in all three. Of the six, four are catenative, and there are no interactions. Table 2 shows these, sorted by an 'overuse' ratio (learners frequency divided by subcorpora mean frequency), thus the first pattern is used 27 times more frequently by the learners than by the *BNC* writers.

Table 2. Verb Patterns Used by Learners but Rare in Subcorpora

Pattern	length	Lrn.n	C1.n	C2.n	C3.n	C.mean	Learner. example	BNC. example	mod	phras	prog	perf	pass	cate-native	Interac-tions
VVB-TO0-VHI	3	18	1	0	1	0.67	want to have	appear to have	-	-	-	-	-	+	27
VVB-TO0-VDI	3	41	0	2	4	2.00	want to do	want to do	-	-	-	-	-	+	20.5
VDB-XX0-VHI	3	39	4	1	2	2.33	do not have	do not have	-	-	-	-	-	-	16.71
VDD-XX0-VHI	3	23	3	3	3	3.00	did not have	did not have	-	-	-	-	-	-	7.67
VVB-TO0-VBI	3	47	6	8	5	6.33	want to be	tend to be	-	-	-	-	-	+	7.42
VDB-XX0-VVI-TO0-VVI	5	28	8	2	5	5.00	do not need to worry	do not need to assume	-	-	-	-	-	+	5.6

At this point, the nuances of the C5 tags are illuminating in their separation of the verbs *be, do* and *have* from lexical verbs. Two patterns are very specific: the sequence VDB-XX0-VHI can only be 'do not have' and similarly, the fourth can only be the past tense form of this, 'did not have.' Except for the last one, these show a striking overuse of the three non-lexical verbs as the second verb in catenative constructions. In all of the three patterns beginning with a lexical verb (i.e., beginning with VVB), *want* is the most frequent verb in the first half, followed by *need*, which together account for around three-quarters of the verbs in this slot.

Before addressing the next question, it may be worthwhile to sample some of the VPs used by learners which were not found significant, if only to give their flavour. The first of these is: {lexical verb} *to* {infinitive verb} {progressive verb}, as in 'want to begin learning.' The pattern is not inherently grammatical. It does occur in the whole *BNC*, just not in the sampled subcorpora (130 times in the written component, though a few of these are mistagged). The second pattern here may stem largely from the lack of one-to-one translation of the Korean word *eopsda* 'to not exist.' The pattern is: *is not* {bare infinitive verb}, as in 'If [it] is not exist [...].' The third pattern is the phrase *have to have* (not followed by another verb). This pattern is again not found in the subcorpora but is present in small quantities in the larger corpus. The word *need* is a more common and economical way to express this, and, looking at the *BNC* instances, *have to have* is frequently followed by a noun denoting official sanctioning (certification, approval), medical needs (injection, a special diet) or personality traits (a good sense of humour), etc. Again, though, these instances were not significant.

There is an important point here that must not be left implicitly stated: there is no repeatedly-used ill-formed VP pattern. Any pattern they use that is not found in the

BNC and which raises a red flag in native speaker's intuitions only occurs once or twice.

4.3. Verb Phrase Patterns Overused by the Learners

The third question is basically a continuation of the previous one as it does not restrict itself to VPs with very low frequency rankings in the subcorpora. The question regards which forms the learners overuse. The data set are those roles indicating (1) presence in all three subcorpora; (2) significance, still at $p<0.001$; (3) mean frequencies in the subcorpora of more than ten; and (4) overuse by the learners (learner frequencies greater than the mean of the subcorpora frequencies). This yields 20 patterns, of which six are one-word VPs and five are two-word patterns. Of these 20, five are catenative, three are modal and three progressive, with only one two-way interaction (catenative and progressive). It is worthwhile to note that the learners did not overuse any VPs that are phrasal, perfective or passive. These instances are shown in Table 3.

Sorted, again, by the overuse ratio, the top two are the phrases *do not have* (used almost 17 times more frequently by the learners) and *did not have* (eight times more frequent). These are followed by five catenative VPs, exemplified by 'want to be,' 'have to pay,' 'come to think,' 'do not need to worry,' and 'prefer living,' plus one more ('want to keep') further down the list.

We also find an overuse of simple, one-word VP patterns. The verb *have* (by itself, as in 'I have an idea') approaches four times the frequency of the *BNC* subcorpora texts. The forms *do* and *did* are both used with around three times the frequency we would expect from native-speaking writers, and *has* is around twice as much. Even one-word lexical verbs are used more than twice for every one occurrence in the subcorpora. The present tense copula (realised as *is, am* and *are*) is slightly overused (1.33 times) but still considered significant.

We also find an overuse of short modal VPs. Modally expressed *do* ('can do,' 'will do,' etc.) occurs with more than three times the expected frequency, while modals with lexical verbs, both negative ('could not explain') and non-negative ('can distinguish'), are used almost twice as often by the learners than are found in the *BNC*-sampled texts.

Table 3. Verb Patterns Overused by Learners at $p<0.001$

Pattern	Length	Lrn.n	C1.n	C2.n	C3.n	C. mean	Learners. example	BNC. example	mod	phras	prog	perf	pass	cate-native	Interac-tions	Overuse. ratio
VDB-XX0-VHI	3	39	4	1	2	2.33	do not have	do not have	-	-	-	-	-	-	0	16.714
VDD-XX0-VHI	3	23	3	3	3	3.00	did not have	did not have	-	-	-	-	-	-	0	7.667
VVB-TO0-VBI	3	47	6	8	5	6.33	want to be	tend to be	-	-	-	-	-	+	1	7.421
VHB-TO0-VVI	3	100	14	9	22	15.00	have to pay	have to find	-	-	-	-	-	-	0	6.667
VVB-TO0-VVI	3	323	66	56	51	57.67	come to think	remember to use	-	-	-	-	-	+	1	5.601
VDB-XX0-VVI-TO0-VVI	5	28	8	2	5	5.00	do not need to worry	do not need to assume	-	-	-	-	-	+	1	5.600
VVB-VVG	2	43	8	10	16	11.33	prefer living	enjoy working	-	-	-	-	-	+	1	3.794
VHB	1	320	71	100	90	87.00	have	have	-	-	-	-	-	-	0	3.678
VM0-VDI	2	76	20	22	26	22.67	can do	can do	+	-	-	-	-	-	1	3.353
VDB	1	100	31	27	36	31.33	do	do	-	-	-	-	-	-	0	3.191
VDD	1	75	28	29	24	27.00	did	did	-	-	-	-	-	-	0	2.778
VHZ	1	182	70	75	85	76.67	has	has	-	-	-	-	-	-	0	2.374
VBB-VVG	2	155	71	74	59	68.00	are going	are appearing	-	-	+	-	-	-	1	2.279
VVB	1	2689	1137	1200	1221	1186.00	live	walk	-	-	-	-	-	-	0	2.267
VBZ-VVG	2	122	58	56	63	59.00	is waiting	is backing	-	-	+	-	-	-	1	2.068
VDB-XX0-VVI	3	162	78	74	92	81.33	do not stay	do not mean	-	-	-	-	-	-	0	1.992
VVD-TO0-VVI	3	142	73	72	80	75.00	want to keep	begin to unfasten	-	-	-	-	-	+	1	1.893
VM0-XX0-VVI	3	197	117	98	102	105.67	could not explain	can not say	+	-	-	-	-	-	1	1.864
VM0-VVI	2	1158	608	653	691	650.67	can distinguish	may present	+	-	-	-	-	-	1	1.780
VBZ	1	1852	1357	1317	1352	1342.00	is	is	-	-	-	-	-	-	0	1.380

VPs in the present tense and progressive aspect are also found to be significantly overused. There are two such patterns, but these differ only in the plurality condition of the subject, and we may consider them a single pattern: 'are going' and 'is waiting.'

4.4. Verb Phrase Patterns Underused by the Learners

The fourth question examines which VP patterns are underused, but still used, by the learners. Our data subset, then, includes those words with are: (1) in all three subcorpora; (2) significantly different; (3) have learner frequencies greater than zero but less than the subcorpora frequency mean. This produces a data set of 23 distinct VPs, of which ten (43%) are passive, eight (35%) are perfective, five are phrasal,

four are modal and three, catenative. There are two three-way interactions and ten two-way interactions, of which three which are both perfective and passive. The learners did not underuse any progressive forms. These are shown in Table 4.

Table 4. Verb Patterns Underused by Learners at $p<0.001$

Pattern	Length	Lrn.n	C1.n	C2.n	C3.n	C.mean	Learners.example	BNC.example	mo	phras	prog	perf	pass	Cate-native	Interac-tions	Under-use.ratio
VHD-VBN-VVN	3	3	34	33	29	**32.00**	had been educated	had been slashed	-	-	-	+	+	-	2	**10.67**
VBD-VVN-TO0-VVI	4	3	31	17	31	**26.33**	was failed to accomplish	was allowed to stand	-	-	-	-	+	+	2	**8.78**
VVB-NNN-AVP	3	2	16	18	17	**17.00**	take a year off	pick the food up	-	+	-	-	-	-	1	**8.50**
VHB-VBN-VVN	3	6	39	47	46	**44.00**	have been lived	have been fascinated	-	-	-	+	+	-	2	**7.33**
VVB-NNN-VVI	3	2	14	14	15	**14.33**	make me feel	let me know	-	-	-	-	-	+	1	**7.17**
VM0-VHI-VBN	3	3	12	22	26	**20.00**	could have been	would have been	+	-	-	+	-	-	2	**6.67**
VHD-VBN	2	7	40	38	35	**37.67**	had been	had been	-	-	-	+	-	-	1	**5.38**
VBD-VVN-AVP	3	5	18	34	24	**25.33**	was freaked out	was sucked out	-	+	-	-	+	-	2	**5.07**
VM0-VHI-VVN	3	11	38	47	47	**44.00**	must have heard	would have found	+	-	-	+	-	-	2	**4.00**
VVD-NNN-AVP	3	5	21	17	22	**20.00**	sent me off	make one up	-	+	-	-	-	-	1	**4.00**
VM0-VBI-VVN	3	56	199	220	220	**213.00**	might be spent	must be planned	+	-	-	-	+	-	2	**3.80**
VVZ-AVP	2	10	35	38	40	**37.67**	goes up	snaps out	-	+	-	-	-	-	1	**3.77**
VM0-XX0-VBI-VVN	4	6	20	23	19	**20.67**	could not be felt	could not be dealt	+	-	-	-	+	-	2	**3.44**
VHZ-VBN-VVN	3	16	60	55	42	**52.33**	has been exported	has been warned	-	-	-	+	+	-	2	**3.27**
VHZ-VVN	2	44	152	144	133	**143.00**	has increased	has proven	-	-	-	+	-	-	1	**3.25**
VBD-VVN	2	123	363	379	387	**376.33**	was attracted	was designed	-	-	-	-	+	-	1	**3.06**
VVZ-TO0-VVI	3	14	40	44	34	**39.33**	tends to treat	needs to provide	-	-	-	-	-	+	1	**2.81**
VHD-VVN	2	68	182	166	186	**178.00**	had studied	had come	-	-	-	+	-	-	1	**2.62**
VBZ-VVN	2	96	228	204	192	**208.00**	is known	is taken	-	-	-	-	+	-	1	**2.17**
VVD-AVP	2	74	131	129	126	**128.67**	found out	galloped back	-	+	-	-	-	-	1	**1.74**
VBB-VVN	2	83	123	145	123	**130.33**	are needed	are based	-	-	-	-	+	-	1	**1.57**
VVZ	1	550	873	870	807	**850.00**	makes	requires	-	-	-	-	-	-	0	**1.55**
VBD	1	1065	1120	1148	1127	**1131.67**	was	was	-	-	-	-	-	-	0	**1.06**

The most severely underused pattern is the past perfect passive, as in 'had been educated,' but as two other perfect passive patterns are found here, we will consider them together, the others being the present perfect passive for first and third person. Combined, these are used an average of 128 times in each subcorpora but only 25 times in the LC, thus 20 times more often by native speakers. Alas, this tells an incomplete story as, if we look at the actual sentences, we see the learners employ these patterns with a low rate of accuracy. Of the 25 uses, only six can be said to correct in terms of passivity (that is, ignoring problems of tense and number, etc.). A recurring problematic word is *change*, which can be passive but requires an agentive force (as in 'rules have been changed' or 'names have been changed' where someone has performed the act of changing) but is used by the learners in contexts like 'the situation has been changed' where 'the situation has changed' is more appropriate. Note that *change* was a lexeme that Cowen, Choi and Kim (2003) also found as a recurring problem in their data of Korean learners. One especially problematic item is found in 'has been happened' as *happen* is not transitive and therefore incapable of being passivised. It would seem a potential trouble spot is in not considering whether there are semantic agents or patients involved in the action, but recall, too, that Ju (2000) used this very example ('happened') in her discussion of conceptualizable agents. Alternatively, there may be formal confusion with the perfect passive ('*have been lived') and the perfect progressive ('have been living') or even the simpler past perfect ('have lived').

Still in the category of significantly underused VP patterns, there are six more passive patterns which are not cast in the perfective aspect. These include past passive catenative phrases ('was designed to maximise'), past passive with adverb particle ('was freaked out'), modal passives, both negative ('could not be felt') and non-negative ('can be heard'), simple present passives ('was designed' and 'is known') and simple past passives ('are needed'). The accuracy rate, however, is notably higher at 86%. Some of these are adjective-like 'statal passives' (Greenbaum and Quirk 1990: 45) such as 'I was disappointed' which may contribute somewhat to the higher accuracy score. Many of the errors, however, seem not to indicate passivity so much as they may be redundantly past-tensed as in 'I was entered the house' and 'birds were floated over the sea' and may indicate problems in tense mastery rather than attempts at passivisation. These could also be interpreted as a less-than-perfect understanding of the perfect aspect, using a form of *be* rather than of *have*, or even just a lack of confidence that single-word verb phrases may be the correct form. But

although all these passive patterns are significantly underused, the most serious problem is clearly the combination of passive and perfective markers.

Another underused patten is past and present VPs with a noun phrase followed by an adverb particle ('took a year off' and 'sent me off'). What's interesting here, however, is when we examine the *BNC* for this pattern, we find a surprising number of instances which are idioms, or at least idiomatic rather than fully compositional. These include *get a move on, keen an eye out, get a kick out, turn the clock back, let the cat out, take the mickey out, take a leaf out, draw the curtain back, take the piss out, put the phone/receiver down* and *take a trip down* (*memory lane*), among others. Thus, an undeveloped sense of idiomaticity may be contributing to the underuse of at least this one pattern.

The results also indicate several phrasal patterns are significantly underproduced by the learners. These patterns include the simplest forms (present tense: 'snaps out', past tense: 'found out') and the interrupted variety in both tense ('take a year off' and 'sent me off'). This is in line with the research discussed above (Dagut and Laufer 1985, Hulstijn and Marchena 1989, Siyanova and Schmitt 2007, among others).

Other forms identified as underproduced include modal perfectives, both copulaic and lexical ('could have been' and 'must have heard'), present tense VPs with noun phrases followed by bare infinitives ('make me feel'), past perfectives with *be* ('had been'), past and present perfectives ('has increased' and 'had studied'), present catenative VPs with infinitives ('tends to treat'), simple present tense lexical verbs in third person ('makes') and the past tense form of the copula ('was').

5. Conclusion

The learners whose writings we examined are university students, meaning they are the final stage of formal education. After graduating, most are unlikely to continue their education in an institutional setting, so this research takes a snapshot of the near-end state of their mastery of English verb phrase patterns.

Some very clear patterns have emerged. Verb phrases in the perfect aspect and passive voice are both found to be particularly underutilised, and when combined, there is an alarming level of inaccuracy. Passives and phrasal verbs are significantly overused. Non-lexical verbs (*do, be* and *have*) are overused, particularly in complex forms where they appear in the non-initial slot, such as in the second half of modal

or catenative phrases.

Further review of these forms at the university level and earlier would no doubt be beneficial. Errors appear to stem from not knowing when to use the passive voice, particularly in free production (hence its underuse), and some uncertainty in how (hence the error rate). This would likely entail a review of verb transitivity. Lessons on which verbs cannot be passivated would also likely benefit the learners in their progress through the mastery of the complex English verb system.

This undertaking has shown that Korean university students are largely making the same verb-related errors as learners of other backgrounds. Other researchers, Koo (2015) for instance, inform us that there may be differences within these similarities. The three largest problem areas (the passive voice, phrasal verbs and perfect aspect) have been heavily researched, although the last of these is often huddled together with tense, which we have not examined (see Alsalmi 2013, Babu and Kumar 2017, Bardovi-Harlig 2000, Shirai 2003). Unmotivated shifts in tense do occur in the LC, but they are not particularly common and are usually momentary. This raises the question of whether tense and aspect should be decoupled as pedagogical topics when learners are cognitively mature enough to make the distinction. Their conflation may overcomplicate the system of encoding time in language.

The errors and the underproduced (and unproduced) forms when examined *en masse* show that the English verb system has not been sufficiently mastered by those in the final stages of formal education in Korea. A call for reform is very much warranted.

References

Alsalmi, M. 2013. Tense and aspect acquisition in L2 English by native speakers of Arabic. *Arab World English Journal* 4(2), 270-282.

Babu1, T. R. and A. P. R. Kumar. 2017. Teaching tenses in a simplest method for non-native speakers. *Journal Of Humanities And Social Science* 22(12), 46-52.

Bardovi-Harlig, K. 2000. *Tense and Aspect in Second Language Acquisition: Form, Meaning, and Use*. Oxford: Blackwell

Benjamin, D. J., J. O. Berger, M. Johannesson, B. A. Nosek, E. J. Wagenmakers, R. Berk, ... and V. E. Johnson. 2018. Redefine statistical significance. *Nature Human Behaviour* 2(1), 6-10.

Bjers, Ma. J. and B. A. Massicotte 2015. Avoidance of the relative clause by Korean

learners of English, *Linguistic Portfolios* 4, Article 4.

Bronshteyn, K. C. and T. Gustafson. 2015. The acquisition of phrasal verbs in L2 English: A literature review. *Linguistic Portfolios* 4, Article 8.

Chen, M. 2013. Overuse or underuse: A corpus study of English phrasal verb use by Chinese, British and American university students. *International Journal of Corpus Linguistics* 18. doi:10.1075/ijcl.18.3.07che

Cowan, R, H. Choi and D. H. Kim. 2003. Four questions for error diagnosis and corrections in CALL. *CALICO Journal* 20(3), 451-463.

Hulstijn, J. H., and E. Marchena. 1989. Avoidance: Semantic or grammatical causes. *Studies in Second Language Acquisition* 11(3), 241-255.

Jahedi, M. and J. Mukundan. 2014. A review on studies of phrasal verb constructions in ESL contexts. *Advances in language and literary studies* 6(1), 157-162.

Ju, M. K. 1997. Overpassivisation errors by second language learners: The effect of conceptualizable agents in discourse. *Studies in Second Language Acquisition* 22, 85-111.

Koo, J. 2015. The avoidance of phrasal verbs: Comparing Korean learners of English with German English learners. *SNU Working Papers in English Linguistics and Language* 13, 165-183.

Perlmutter, D. M. 1978. Impersonal passives and the unaccusative hypothesis. *Proceedings of the Annual Meeting of the Berkeley Linguistics Society* 38, 157-189.

Shirai, Y. 2003. The acquisition of tense-aspect morphology and the regular-irregular debate. *ZAS Papers in Linguistics* 29, 195-209.

Siyanova, A. and N. Schmitt. 2007. Native and nonnative use of multi-word vs. one-word verbs. *IRAL* 45, 119-139.

Zobl, H. 1989. Canonical structures and ergativity. In S. M. Gass and J. Schachter, eds., *Linguistic Perspectives on Second Language Acquisition,* 203-221. New York: Cambridge University Press.

Examples in: English
Applicable Languages: English
Applicable Level: Tertiary

Kevin Parent
Professor, Korea Maritime and Ocean University

Department of English Language and Literature
727 Taejeong-ro, Dongsam-dong, Youngdo-gu
Busan, South Korea
Tel: 051) 410-4593
E-mail: parent@kmou.ac.kr

Received: November 30, 2018
Revised: December 12, 2018
Accepted: December 16, 2018