

〔論文〕

Examining targeted phrasal verb items and their stems in multiple-choice vocabulary questions using authenticity and frequency in general corpora

Hiroko USAMI*

1. Introduction

A corpus is defined as an electronic collection of written or spoken texts that are selected to be representative of a particular language or genre for linguistic analysis (e.g., Baker et al., 2006; McEnery et al., 2006). Based on the idea that corpus linguistics is a methodology, corpora have long been applied to various linguistic fields, such as stylistics, pragmatics, sociolinguistics, translation studies, forensic linguistics (e.g., Andersen, 2010; Cotterill, 2010). Language teaching is also an area to which corpora have long been applied, including computer-assisted language learning (CALL), data-driven learning (DDL), and compiling teaching materials and dictionaries (e.g., Hunston, 2002; Chambers, 2010; Cheng, 2010; Walsh, 2010).

The area of language testing and assessment is a relatively new field of the corpus application (e.g., Hunston, 2002). Recently, various methods of applying different types of corpora, such as general, learner, and specialised corpora, to language testing and assessment have been increasingly recognised and discussed (e.g., Park, 2014; Cushing, 2017). For example, general corpora and specialised corpora containing academic, business, and other field-specific English can play an important role in writing, editing, or proofreading items by investigating vocabulary or grammar item frequency, authenticity, word senses, and appropriate distractors presented in authentic texts of corpora (Ball, 2001; Barker, 2006; Crossley et al., 2007; Jaén, 2007; Hughes, 2008). Furthermore, learner corpora (e.g., the Cambridge Learner Corpus, the Cambridge Corpus of Spoken and Learner English) can be utilised for standardisation across and within different exam levels and suites, the English Profile Programme (EPP) for the purpose of characterising learners' English across proficiency levels, test development and validation, and automated evaluation of learners' output (Boyle & Booth, 2000; Ball, 2001; Barker, 2004). However, few specialised corpora containing actual past exam questions have been compiled; such specialised corpora have not been applied to further investigate frequently targeted vocabulary in tests and general or learner corpora have not been applied to examine the English used in examinations adequately yet.

Therefore, the aim of this paper is to discuss how the Exam Corpus, a specialised corpus created by the author, can be applied to language testing and assessment. This will help better inform and examine the English presented in multiple-choice vocabulary questions from English proficiency tests using different types of corpora. Among the multiple-choice vocabulary questions in the Exam Corpus, this study specifically analysed 1) the frequently targeted phrasal verb items and 2) the stems presented in the frequently targeted phrasal verb items in examinations.

投稿日 2022 年 11 月 30 日 受理日 2023 年 1 月 25 日

*教養学部准教授

2. Literature Review

2 – 1 Application of Corpora to Language Testing and Assessment

Among the various applications of corpora to language teaching, the area of language testing and assessment has been regarded as a relatively new field (e.g., Hunston, 2002). Recently, various methods to apply different types of corpora to language testing and assessment have been increasingly discussed in the journal of *Language Testing* (e.g., Park, 2014; Cushing, 2017). Following theoretical and practical suggestions from various researchers (Alderson, 1996; Barker, 2006), how different types of corpora can be utilised in language testing and assessment has been extensively discussed from a different perspective.

First, general corpora, containing general English such as American or British English, have been utilised in various ways (Ball, 2001; Barker, 2006; 2010), such as writing items by obtaining and examining authentic texts and editing or proofreading items by checking the frequency and authenticity of their collocations, word senses in contexts, and appropriate distractors presented in authentic texts of corpora (e.g., Crossley et al., 2007; Jaén, 2007; Hughes, 2008).

Second, learner corpora containing test-takers' written and spoken output collected from examinations (e.g., the Cambridge Learner Corpus, the Cambridge Corpus of Spoken and Learner English, and so on) have been compiled (Boyle & Booth, 2000; Ball, 2001; Barker, 2010) and utilised in various ways: standardisation across and within different exam levels and suites, characterisation of learners' language, including collocational patterns and errors at different proficiency levels in the framework of the English Profile Programme (EPP) (Alexopoulou, 2008), test development, including compiling word lists, validation purposes, and automatically scoring of learners' written and spoken productions (Burstein et al., 2004).

Finally, specialised corpora, containing particular types of texts, as well as general and learner corpora, can play a role in language testing and assessment. As learner corpora, the most basic way to apply corpus approaches is to actually compile specialised corpora suitable for each research purpose. They have been applied in various ways, including a listening test and pronunciation assessment with the Michigan Corpus of Academic Spoken English (MICASE) (Read, 2002; Levis & Cortes, 2008), and the development of a grammar test with the British Academic Written English corpus (Sharpling, 2010). Usami (2021) compiled the Exam Corpus containing various English proficiency tests and suggested possible applications through which the English presented in the stems, answers, and distractors, and their CEFR levels and collocations can be examined across different English proficiency tests.

2 – 2 Corpus Applications to Vocabulary Tests

Corpora have been utilised and analysed specifically in vocabulary tests in various ways as well. First, existing examinations have been investigated and validated by compiling specialised corpora and using different types of corpora. A specialised corpus, the TOEFL 2000 Spoken and Written Academic Language Corpus (T2K-SWAL), was developed in order to examine university-level language skills and provide an alternative to the intuitions of TOEFL test item writers (Biber et al., 2001; 2004; Biber, 2006). Similarly, a specialised corpus constructed by the author, the University Entrance Examinations in Japan (UEEJ) Corpus (i.e., the previous version of the Exam Corpus), containing multiple-choice vocabulary and grammar questions from Japanese university entrance examinations, was developed in order to examine and analyse the English presented in the examinations. The collocations used in the stems and frequently targeted grammar, answers, and distractors were focused on and analysed using learners' productions in the Longman Learners' Corpus and authentic English in the British National Corpus (BNC) (Usami, 2005; 2012; 2015).

As a relatively challenging approach to utilise corpora for vocabulary tests, wordlists compiled based on specialised or general corpora have been constructed and employed for writing items. For example, Business

English Texts Corpus, a web-based collection of business texts, was constructed by UCLES and utilised to develop the Business English Certificate (BEC) Preliminary wordlist. Item writers can create realistic examination tasks at specific proficiency levels, indicating the collocational patterns of certain words or phrases and testing different senses of words in real texts (Ball, 2001; 2002; Horner & Strutt, 2004).

This is perhaps the most challenging way to apply corpora to language testing and assessment. Various corpora have been applied to generally create and develop vocabulary tests, such as cloze tests (Rees, 1998), an nth-deletion cloze test (Coniam, 1997), a blank-filling vocabulary test (Shillaw, 1994), receptive and productive vocabulary tests (Merino, 2000), and receptive and productive collocation tests (Jaén, 2007). Usami (2012; 2015) created and improved the stems and answers of multiple-choice vocabulary and grammar questions by checking authentic English from the BNC, and effective distractors using the learners' errors from the Longman Learners' Corpus.

2 – 3 Authenticity and Frequency in Language Testing

In the application of corpus linguistics to language teaching, authenticity and frequency information obtained in general corpora have attained an important role in compiling and examining teaching materials, investigating frequently targeted grammar and vocabulary items, and the English used in test texts, stems, answers, and distractors in multiple-choice questions (e.g., Shortall, 2007). Most researchers agree that authentic English in general corpora should be used in language teaching (e.g., Hunston & Francis, 1999). However, certain researchers disagree because authentic English is not always necessary for language teaching and there are other factors such as teachability that can be more important than authentic English (e.g., Barlow, 1996; Cook, 1998; Widdowson, 2000).

On the other hand, analyses of authenticity and frequency information obtained from general corpora have not been discussed adequately in terms of language testing, especially the English presented in examinations, in spite of the fact that analyses of authenticity and frequency information have been discussed from the perspective of teaching materials. In language testing and assessment, construct validity along with reliability has been recognised as important, which might be the reason why corpora have been regarded as difficult to apply to language testing (e.g., Alderson, 1996). However, authentic texts in general corpora should be applied to language testing more in order to validate construct validity. In English proficiency tests, apart from the ones categorised across proficiency levels, targeted grammatical categories and vocabulary are already learned overall. If a particular grammatical category or vocabulary is tested or specific stems are presented many times in English proficiency tests, test takers might perceive them as important because the target grammatical categories or vocabulary are presented or used more in actual English.

3. Research Questions

The above literature review establishes that different types of corpora have been employed in language testing and assessment, especially in terms of vocabulary tests. However, there is a dearth of research in this regard; few specialised corpora containing actual past English proficiency tests have been compiled, little work applying such specialised corpora to investigate and analyse frequently targeted vocabulary tested has been carried out, and general or learner corpora have not been applied to examine the English used in examinations adequately. Therefore, the aim of this paper is to discuss the application of a specialised corpus created by the author, the Exam Corpus, to language testing and assessment for the purpose of informing and examining the English presented in multiple-choice vocabulary questions from English proficiency tests using different types of corpora. Among the vocabulary questions, this study specifically focused on phrasal verbs and addressed the following research questions (RQs):

RQ1: What is the most frequently targeted phrasal verb item in English proficiency tests?

RQ2: What is the frequency of the frequently targeted phrasal verb items in general corpora?

RQ3: What are the most frequent stems presented in the most frequently targeted phrasal verb item in English proficiency tests?

RQ4: What are the frequent collocations or phrases presented in the stems of the most frequently targeted phrasal verb item in a general corpus?

4. Method

4 – 1 The Exam Corpus

In the course of this research, a specialised corpus, the Exam Corpus, was analysed to examine the most frequently targeted phrasal verb item in English proficiency tests, and the most frequent stems presented in the most frequently targeted phrasal verb item in English proficiency tests, thereby answering RQs 1 and 3, respectively.

The Exam Corpus is a specialised corpus created by the author in order to be applied to language testing and assessment and examine existing English proficiency tests. Table 1 shows the number of files, that is, questions, of the vocabulary questions across English proficiency tests contained in the Exam Corpus (Ver. 1.2).

Table 1

The number of vocabulary question files across English proficiency tests in the Exam Corpus (Ver. 1.2)

English proficiency tests	UEEJ	TEAP	EIKEN	TOEFL	TOEIC	CMB	Total
# of files/questions	11,860	60	1,670	0	423	380	14,393

Note. UEEJ = University Entrance Examinations in Japan. TEAP = Test of English for Academic Purposes. CMB = Cambridge English.

The Exam Corpus (Ver. 1.2) contains 934,682 words presented in 23,365 exclusively multiple-choice vocabulary and grammar questions from four kinds of English proficiency tests, taken from only the official test books published after 2000 by authorised examination boards or publishers: 1) University Entrance Examinations in Japan (UEEJ) for 2001–2022, 2) English proficiency examinations administered in Japan (Test of English for Academic Purposes (TEAP), EIKEN—pre-1st, 2nd, pre-2nd, and 3rd grades for 2011–2019), 3) English proficiency examinations created by the Educational Testing Service (ETS) in the U.S. (Test of English as a Foreign Language (TOEFL) in 2012, Test of English for International Communication (TOEIC) for 2005–2020); and 4) English proficiency examinations created by Cambridge Assessment English in the U.K. (Cambridge English–Key English Test (KET), Preliminary English Test (PET), First Certificate in English (FCE) for 2006–2014). In this study, only vocabulary questions (both synonym and gap-filling questions, as follows) totalling up to 14,393 questions were analysed, most of which (i.e., 11,860 questions) were presented in university entrance examinations in Japan.

Synonym question:

The primary purpose of his visit is to improve trading relations.

A) further B) solitary C) political D) main

Gap-filling question:

Please () your papers to me by the end of the month.

A) hand out B) hand in C) hand down D) hand of

(Usami, 2018: 37)

4 – 2 General Corpora

Besides the Exam Corpus, general corpora were analysed to examine the frequency of the frequently targeted phrasal verb items in general corpora and frequent collocations or phrases presented in the stems of the most frequently targeted phrasal verb item in a general corpus, thereby answering RQs 2 and 4.

In order to examine general English, two kinds of general corpora were analysed: the British National Corpus (BNC) for approximately 100 million words (90% for written and 10% for spoken texts) of British English collected around in 1990 with the BNCweb, and Corpus of Contemporary American English (COCA) for approximately one billion words (80% for written and 20% for spoken texts) of contemporary American English on the website (<https://www.english-corpora.org/coca/>). These general corpora are varied in terms of the time constructed, size, and kind of English, though the ratio of written to spoken texts was almost the same (approximately 80-90% for written and approximately 10-20% for spoken texts).

4 – 3 #LancsBox

To analyse vocabulary questions in the Exam Corpus, a software package, #LancsBox (see <http://corpora.lancs.ac.uk/lancsbox/>) developed at Lancaster University was used. Among the functions available in #LancsBox, the functions called *Words* (for wordlists), *Ngrams* (for n-grams), and *KWIC* (for concordances) were utilised in this research.

5. Analysis and Discussion

5 – 1 Frequency of Targeted Phrasal Verb Questions in the Exam Corpus

First, in this section, RQ 1 (i.e., What is the most frequently targeted phrasal verb item in English proficiency tests?) was examined. The Exam Corpus was analysed using #LancsBox to obtain frequently targeted phrasal verb items.

Table 2 shows the top 20 most frequently targeted phrasal verbs in English proficiency tests. The phrasal verb *put up with* was the most frequently targeted in the tests. Actually, out of 36 questions targeting the phrasal verb *put up with* (half for synonym questions and another half for gap-filling questions), all questions were presented in Japanese university entrance examinations, apart from one question presented in EIKEN 2nd grade. These 35 questions were tested in 25 different universities for 15 different years.

The second and third most frequently targeted phrasal verbs were *turn down* and *get over*, respectively, and overall, the frequency gradually and evenly decreased. Furthermore, most of the phrasal verbs appeared to contain basic verbs (e.g., *put*, *turn*, *get*, *call*, *come*, *take*, *make*, *look*, etc.) followed by particles and/or adverbs (e.g., *up*, *down*, *over*, *out*, *for*, *off*, *with*, *out*, *to*, *from*, etc.).

Table 2

Frequency of the targeted phrasal verb items in English proficiency tests

Rank	Phrasal verb	Frequency	Rank	Phrasal verb	Frequency
1	put up with	36	11	call for	22
2	turn down	34	11	make up for	22
3	get over	33	13	carry out	21
4	figure out	30	13	do away with	21
5	stand for	28	13	make out	21
6	call for	26	13	take off	21
6	put off	26	17	catch up with	20
8	account for	25	17	look after	20
9	come up with	24	19	bring up	19
10	take after	23	19	look up to	19

5 – 2 Frequency of Targeted Phrasal Verb Items in General Corpora

In this section, RQ 2 (i.e., What is the frequency of the frequently targeted phrasal verb items in general corpora?) was examined, compared to the frequency in the Exam Corpus. The general corpora, such as the BNC and COCA, were analysed to examine the frequency of the targeted phrasal verb items in the general corpora.

Table 3

Frequency (per million) of the targeted phrasal verb items in the general corpora

Corpora	Exam Corpus		BNC		COCA	
Phrasal verb	Rank	Freq.	Rank	Freq. (per million)	Rank	Freq. (per million)
put up with	1	36	7	6.7	8	5.4
turn down	2	34	5	10.2	7	7.9
get over	3	33	6	7.9	5	15.6
figure out	4	30	9	2.1	1	57.0
stand for	5	28	4	13.2	6	14.9
call for	6	26	1	49.6	2	56.3
put off	6	26	8	5.6	9	4.0
account for	8	25	2	43.8	4	33.6
come up with	9	24	3	19.2	3	40.1
take after	10	23	10	1.9	10	1.2

Table 3 shows the frequency (per million) of the top 10 most frequently targeted phrasal verb items shown in Table 2 in the general corpora. The lemma forms (e.g., *do*, *does*, *did*, *done*, *doing* of the lemma *DO*) of these phrasal verbs were counted per million in the general corpora. As shown in Table 3, the phrasal verb *PUT up with* was not very frequently presented in the general corpora, though it was the most frequently tested in the Exam Corpus. In fact, the phrasal verb *PUT up with* was presented in 6.7 and 5.4 times per million, ranking the seventh and eighth, in the BNC and COCA, respectively, which is relatively low.

On the other hand, there are some discrepancies in frequencies and ranks between the Exam Corpus

and general corpora. The phrasal verb *CALL for* was the most and the second most frequently presented in the general corpora and ranked the first and second in the BNC and COCA, respectively, though it ranked the sixth in the Exam Corpus. In addition, the phrasal verb *ACCOUNT for* was the second most frequently presented in British English in the BNC and relatively frequent in American English in the COCA, ranking the fourth. However, it was much less frequently tested in the Exam Corpus, ranking the eighth. Similarly, the phrasal verb *COME up with* was relatively frequently presented in the general corpora, ranking the third for both, whereas it was the second least frequently tested in the Exam Corpus, ranking the ninth.

As noted above, there are some discrepancies in frequencies and ranks between the Exam Corpus and general corpora and even between the general corpora of the BNC and COCA. The phrasal verb *FIGURE out* was the second least frequent in the BNC, ranking the ninth, whereas it was the most frequently presented in the COCA. However, in the Exam Corpus, it was relatively frequent and ranked the fourth.

On the other hand, interestingly, there are similarities between the Exam Corpus and general corpora; the phrasal verb *TAKE after* is the least frequently presented both in the general corpora and Exam Corpus.

5 – 3 Stems in *Put up with* in the Exam Corpus

In this section, RQ 3 (i.e., What are the most frequent stems presented in the most frequently targeted phrasal verb item in English proficiency tests?) was examined. The Exam Corpus was analysed using #LancsBox to obtain the frequent stems presented in the most frequently targeted phrasal verb item, that is, *put up with*.

Table 4 shows the wordlist, 2-grams, and 3-grams presented in the stems of the question item *put up with*. As shown in Table 4, some patterns of the words or phrases presented can be observed. First, personal pronouns of the nominative case, such as *I*, *she*, and *we*, may have been used as the subject of *put up with*. In Table 4, *I* (17 times), *she* (eight times), and *we* (four times) were shown in the wordlist, *I can't* (eight times), *I cannot* (four times), and *anymore I* (twice) in the 2-grams, and *I cannot put* (four times), *I can't zzz* (three times), *any longer I* (twice), *I can't put* (twice), *I can't zzz* (twice), *longer I cannot* (twice), *she couldn't put* (twice), and *she refused to* (twice) in the 3-grams. Actually, in most cases, it was the subject *I* that *put up with* something, as in (1).

(1) I can't put up with children shouting while I'm studying. (Obunsha, 2003: 76)

Second, some modal verbs were shown; *can't* (nine times), *cannot* (seven times), *can* (four times), *couldn't* (four times), and *have* (four times) were shown in the wordlist, *I can't* (eight times), *cannot put* (six times), *couldn't put* (four times), *I cannot* (four times), *can't zzz* (three times), and *have to* (three times) in the 2-grams, and *cannot put up* (six times), *I cannot put* (four times), *can't put up* (four times), *couldn't put up* (three times), *I can't zzz* (three times), *can't zzz up* (twice), *have to put* (twice), *I can't put* (twice), *I can't zzz* (twice), *longer I cannot* (twice), *must put up* (twice), and *she couldn't put* (twice) in the 3-grams. Actually, in most cases, the phrasal verb *put up with* was accompanied by these modal verbs, especially *can* and negative forms *can't* or *cannot*, as follows:

(2) I can't put up with the noise any longer. (Obunsha, 2002: 348)

(3) The host could not possibly put up zzz her attitude. (Obunsha, 2003: 55)

Furthermore, the modal-like expression *had no choice but to* was also used as follows:

(4) I had no choice but to put up with my client's rude behavior.

(Obunsha, 2021: 256)

Table 4

Wordlist, 2-grams, and 3-grams presented in the stems of *put up with* in the Exam Corpus

Rank	word	Freq.	Rank	2-gram	Freq.	Rank	3-gram	Freq.
1	up	26	1	up with	23	1	put up with	18
2	put	25	2	put up	21	2	up with the	10
2	with	25	3	with the	11	3	cannot put up	6
4	the	20	4	I can't	8	4	noise any longer	5
5	zzz	18	5	any longer	7	4	up with his	5
6	I	17	6	cannot put	6	4	zzz up with	5
7	to	12	7	noise any	5	7	I cannot put	4
8	any	11	7	the noise	5	7	up with this	4
9	his	9	7	to put	5	7	with the noise	4
9	can't	9	7	with his	5	7	can't put up	4
11	she	8	7	zzz up	5	11	couldn't put up	3
12	cannot	7	12	any more	4	11	I can't zzz	3
12	longer	7	12	couldn't put	4	11	put up zzz	3
12	noise	7	12	I cannot	4	11	the noise any	3
15	behavior	5	12	put zzz	4	11	to put zzz	3
16	can	4	12	with this	4	16	any longer I	2
16	couldn't	4	17	can't zzz	3	16	behavior any more	2
16	have	4	17	have to	3	16	can't zzz up	2
16	more	4	17	up zzz	3	16	have to put	2
16	of	4	17	zzz her	3	16	I can't put	2
16	this	4	21	anymore I	2	16	I can't zzz	2
16	we	4				16	longer I cannot	2
						16	must put up	2
						16	put zzz with	2
						16	she couldn't put	2
						16	she refused to	2
						16	this noise any	2
						16	to put up	2
						16	with his bad	2
						16	with this noise	2

Note. The symbol *zzz* is replaced with a parenthesis.

Third, the words and phrases consisting of specific adverbial expressions were also found in Table 4; *any* (11 times), *longer* (seven times), and *more* (four times) in the wordlist, *any longer* (seven times), *noise any* (five times), *any more* (four times), and *anymore I* (twice) in the 2-grams, and *noise any longer* (five times), *the noise any* (three times), *any longer I* (twice), *behavior any more* (twice), *longer I cannot* (twice), and *this noise any* (twice) in the 3-grams. These words were collocated as *any longer* and *any more*. Actually, these adverbial expressions were collocated with *put up with* as follows:

(5) I can't put up with your rudeness any more. (Obunsha, 2005: 119)

(6) She cannot put up with the noise any longer. (Obunsha, 2012: 22)

Fourth, the words and phrases that were probably used as the object of *put up with*, that is, the subject someone puts up with, were also presented: *noise* (seven times) and *behavior* (five times) in the wordlist, *noise any* (five times) and *the noise* (five times) in the 2-grams, and *noise any longer* (five times), *with the noise* (four times), *the noise any* (three times), *behavior any more* (twice), *this noise any* (twice), and *with this noise* (twice) in the 3-grams. Furthermore, the determiners or possessive pronouns specifying the object of *put up with* were also found; *the* (20 times), *his* (nine times), and *this* (four times) in the wordlist, *with the* (11 times), *the noise* (five times), *with his* (five times), *with this* (four times), and *zzz her* (three times) in the 2-grams, and *up with the* (10 times), *up with his* (five times), *up with this* (four times), *with the noise* (four times), *the noise any* (three times), *this noise any* (twice), *with his bad* (twice), and *with this noise* (twice) in the 3-grams. Actually, in most cases, the determiners or possessive pronouns followed by objects are presented as follows:

(7) I can't put up with his childish behavior any more. (Obunsha, 2011: 291)

(8) I cannot put up with this noise any longer. (Obunsha, 2002: 348)

5 – 4 Frequent Collocations or Phrases in *Put up with* in the BNC

Finally, in this section, RQ 4 (i.e., What are the frequent collocations or phrases presented in the stems of the most frequently targeted phrasal verb item in a general corpus?) was examined. For a general corpus, only the BNC was analysed in order to obtain frequent collocations or phrases in a general corpus. The COCA had to be excluded because the collocation in the COCA is calculated with different statistics from that of the Exam Corpus and BNC, even though the American English in the COCA would be suitable in the context of English instruction in Japan.

Table 5

Collocates (3 left – 3 right) of *PUT up with* in the Exam Corpus and BNC

Rank	Exam Corpus	BNC	Rank	Exam Corpus	BNC
1	I	to	11	behavior	why
2	the	have	12		nonsense
3	his	prepared	13		could
4	any	had	14		'll
5	cannot	it	15		I
6	noise	just	16		willing
7	to	you	17		longer
8	can't	lot	18		can
9	longer	n't	19		me
10	she	having	20		they

Table 5 shows the collocates of 3 left to 3 right of *PUT up with* in the Exam Corpus and BNC using the log-likelihood value. As shown in Table 5, some of the words collocated with *PUT up with* were shared with those in the Exam Corpus; pronouns (e.g., *I*) and the word consisting of adverbial phrase (e.g., *longer*).

However, interestingly, some words that were not presented in the stems of *put up with* in the Exam

Corpus were actually presented in the BNC. As for the words occurring before *PUT up with*, the word *just* occurred just before *PUT up with* (e.g., *just put up with*) or just before *have to* or *had to* (e.g., *just have to*, *just had to*) as follows:

(9) You will just have to put up with it. (BNC: File No. BN0 379)

Furthermore, the words *prepared* and *willing* occurred before *PUT up with* accompanied with *to* (e.g., *prepared to put up with*, *willing to put up with*) as follows:

(10) She's obviously prepared to put up with your terms. (BNC: File No. H8S 4135)

(11) Dentists won't be willing to put up with such a pay cut. (BNC: File No. K1F 1370)

As well as affirmative sentences, *PUT up with* was used in interrogative sentences accompanied with the word *why* as follows:

(12) So why do they put up with it? (BNC: File No. HR7 1058)

On the other hand, the words occurring after *PUT up with* were also found. The word *it* occurred directly after *PUT up with* (e.g., *put up with it*) as follows:

(13) People would never put up with it before then. (BNC: File No. A0N 1407)

Moreover, the word *lot* occurred both before and after *PUT up with* accompanied by *a* (e.g., *a lot to put up with*, *put up with a lot*) as follows:

(14) She's had a lot to put up with. (BNC: File No. KC5 2623)

(15) Neither party will put up with a lot. (BNC: File No. KGP 972)

Examining the collocates according to their part of speech in more detail, some similar or different frequencies can be observed. Table 6 shows the pronouns and modals collocates of 3 left to 1 left of *PUT up with* in the Exam Corpus and BNC. According to Table 6, the pronouns frequent only in the BNC are *you*, *they*, *we*, and *he*, which is slightly different from the frequency obtained from the Exam Corpus. In the Exam Corpus, the pronoun *I* was by far the most frequent, whereas other pronouns *you* and *they* were not presented in the stems. In addition, the modal lemma WILL (e.g., *will*, *ll*, *would*) were also frequent as well as the lemma HAVE (e.g., *have*, *had*, *having*, *'ve*) and the lemma CAN (e.g., *could*, *can*) in the BNC, such as (16) and (17). In the Exam Corpus, the modal WILL was not presented, and instead, the lemma CAN (e.g., *could*, *can*) was much more frequent sometimes occurred with *not*.

(16) You wouldn't believe what I have to put up with. (BNC: File No. AC3 2565)

(17) The symptoms that he had put up with for 20 years now became a real problem. (BNC: File No. BM1 885)

Table 6

Pronoun and modal collocates (3 left – 1 left) of *PUT up with* in the Exam Corpus and BNC

Pronouns			Modal		
Rank	Exam Corpus	BNC	Rank	Exam Corpus	BNC
1	I	you	1	cannot	have
2	she	I	2	can't	had
3		they	3		could
4		she	4		having
5		who	5		can
6		we	6		'll
7		one	7		will
8		he	8		would
9		it	9		've
10			10		got

Table 7

Adverbial expressions and objects (1 right – 3 right) of *PUT up with* in the Exam Corpus and BNC

Adverbial expressions			Object		
Rank	Exam Corpus	BNC	Rank	Exam Corpus	BNC
1	any	anything	1	the	it
2		any	2	his	nonsense
3		more	3	noise	this
			4		me
			5		such
			6		for
			7		lot
			8		.
			9		all
			10		sort

Table 7 shows the adverbial collocates of 1 right to 3 right of *PUT up with* in the Exam Corpus and BNC. According to the Table 7, the words *any* and *more* were used though they were not so frequent in the BNC, such as (18) and (19). Instead, in the Exam Corpus, the words *longer* and *more* were used, but *anything* was not presented.

(18) Ellcock has put up with more adversity than most people expect in a lifetime. (BNC: File No. BN9 892)

(19) And they'll put up with any conditions. (BNC: File No. HLU 540)

Finally, the pronoun *it* was frequently used in the BNC, whereas the determiner or possessive pronouns *the*, *his*, and *this* were frequently used in the Exam Corpus. Furthermore, it is a concrete noun such as *nonsense*, *situation* (rank 14), and *conditions* (rank 18) that are/were put up with in the BNC, while *noise* and *behavior* were frequently used in the Exam Corpus as follows:

(20) People would never put up with it before then.

(BNC: File No. A0N 1407)

(21) Of course you wouldn't put up with such nonsense.

(BNC: File No. AS3 1464)

6. Conclusion

This paper analysed and discussed how the Exam Corpus can be utilised in language testing and assessment in order to better inform and examine the English presented in multiple-choice vocabulary questions from English proficiency tests using different types of corpora. Specifically, phrasal verb questions in the multiple-choice vocabulary questions from English proficiency tests were analysed in terms of 1) the frequently targeted phrasal verb items, 2) the frequency of the frequently targeted phrasal verb items in general corpora, 3) the frequent stems presented in the most frequently targeted phrasal verb item in English proficiency tests, and 4) the frequent collocations or phrases presented in the stems of the most frequently targeted phrasal verb item in a general corpus, for each RQ.

First, out of the 20 most frequently targeted phrasal verbs, the phrasal verb *put up with* was the most frequently tested in English proficiency tests.

Second, regarding the frequency of the frequently targeted phrasal verb items in general corpora, the phrasal verb *PUT up with* was not so frequently presented in the general corpora, though it was the most frequently tested in the Exam Corpus. Instead, the phrasal verbs *CALL for*, *ACCOUNT for*, and *COME up with* were frequently presented in the general corpora, though they were less frequently tested in the Exam Corpus. Furthermore, the phrasal verb *TAKE after* was the least frequently presented both in the general corpora and Exam Corpus.

Third, regarding the frequent stems presented in the most frequently targeted phrasal verb item, some patterns of the words or phrases were observed for the stems presented in the questions. Specific pronouns such as *I*, *she*, and *we* were used as the subject of *put up with* in the stems. The modal verbs *can't*, *cannot*, *can*, *couldn't*, and *have* were followed by the target *put up with*. Moreover, the adverbial expressions *any longer* and *any more* were presented in the target. Some specific patterns were also seen in the subject, that is, the determiner or possessive pronoun, such as *the*, *his*, and *this*, followed by the subject someone puts up with, such as *noise* and *behavior*.

Fourth, concerning the frequent collocations or phrases in the stems of the most frequently targeted item in the general corpus, some of the collocates with *put up with* were shared with those in the Exam Corpus; pronouns (e.g., *I*) and the word consisting of adverbial phrase (e.g., *longer*). However, interestingly, some words that were not presented in the stems of *put up with* in the Exam Corpus were actually presented in the BNC; specifically: *just*, *prepared*, *willing*, *why*, *it*, and *lot*.

In this paper, the most frequently targeted phrasal verb item in the Exam Corpus *put up with* was examined. However, in the future, the targeted skills, kinds of English proficiency tests, and kinds of targeted vocabulary items can be expanded to be analysed and compared to those in general corpora and other targeted skills (e.g., grammar, listening, conversation, reading) and other English proficiency tests (e.g., TOEIC, EIKEN). Moreover, why some patterns or fixed expressions have been used and presented repeatedly in stems of multiple-choice vocabulary questions in English proficiency tests can be examined and discussed. This is likely attributable to the facts that item writers do not generally refer to authentic English in general corpora, but past exam questions or reference books, and that item writers use their intuitions or their first language's fixed expressions.

Acknowledgements

This research was funded by a Japan Society for the Promotion of Science (JSPS) KAKENHI grant (20K00814). I would like to thank Editage (www.editage.com) for English language editing.

An earlier version of this paper was presented at the Language Testing Forum (LTF) at Lancaster University in 2021.

References

- Alderson, J. C. (1996). Do corpora have a role in language assessment? In J. Thomas & M. Short (Eds.), *Using corpora for language research: Studies in the honour of Geoffrey Leech* (pp. 248-259). Longman.
- Alexopoulou, T. (2008). Building new corpora for English Profile. *Research Notes*, 33, 15-19.
- Andersen, G. (2010). How to use corpus linguistics in sociolinguistics. In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics* (pp. 547-562). Routledge.
- Baker, P., Hardie, A., & McNery, T. (2006). *A glossary of corpus linguistics*. Edinburgh University Press.
- Ball, F. (2001). Using corpora in language testing. *Research Notes*, 6, 6-8.
- Ball, F. (2002). Developing wordlists for BEC. *Research Notes*, 8, 10-13.
- Barker, F. (2004). Using corpora in language testing. *Modern English Teacher*, 13(2), 65-67.
- Barker, F. (2006). Corpora and language assessment: Trends and prospects. *Research Notes*, 26, 2-4.
- Barker, F. (2010). How can corpora be used in language testing? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics* (pp. 633-645). Routledge.
- Barlow, M. (1996). Corpora for theory and practice. *International Journal of Corpus Linguistics*, 1, 1-37.
- Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. John Benjamins.
- Biber, D., Reppen, R., Clark, V., & Walter, J. (2001). Representing spoken language in university settings: The design and construction of the spoken component of the T2K-SWAL Corpus. In R. C. Simpson & J. M. Swales (Eds.), *Corpus linguistics in North America. Selections from the 1999 symposium* (pp. 48-57). The University of Michigan Press.
- Biber, D., Conrad, S.M., Reppen, R., Byrd, P., Helt, M., Clark, V., Cortes, V., Csomay, E., & Urzua, A. (2004). *Representing language use in the university: analysis of the TOEFL 2000 spoken and written academic language corpus*. TOEFL Monograph Series No. 25. ETS.
- Boyle, A. & Booth, D. (2000). The UCLES/CUP Learner Corpus. *Research Notes*, 1, 10.
- Burstein, H., Chodorow, M., & Leacock, C. (2004). Automated essay evaluation: The criterion online writing evaluation service. *AI Magazine*, 25(3), 27-36.
- Chambers, S. (2010). What is data-driven learning? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics* (pp. 345-358). Routledge.
- Cheng, W. (2010). What can a corpus tell us about language teaching? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics* (pp. 319-332). Routledge.
- Coniam, D. (1997). A preliminary inquiry into using corpus word frequency data in the automatic generation of English language cloze tests. *CALICO Journal*, 16(2-4), 15-33.
- Cook, G. (1998). The uses of reality: a reply to Ronald Carter. *ELTJ*, 52, 57-64.
- Cotterill, J. (2010). How to use corpus linguistics in forensic linguistics. In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge Handbook of Corpus Linguistics* (pp. 578-590). Routledge.
- Crossley, S. A., Louwerse, M., McCarthy, P. M., & McNamara, D. S. (2007). A linguistic analysis of simplified and authentic texts. *Modern Language Journal*, 91, 15-30.
- Cushing, S. T. (2017). Corpus linguistics in language testing research. *Language Testing*, 34(4), 441-449.
- Horner, D., & Strutt, P. (2004). Analysing domain-specific lexical categories: evidence from the BEC written corpus. *Research Notes*, 15, 6-8.

- Hughes, G. (2008). Text organisation features in an FCE reading gapped sentence task. *Research Notes*, 31, 26-31.
- Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge University Press.
- Hunston, S., & Francis, G. (1999) *Pattern Grammar: A corpus-driven approach to the lexical grammar of English*. Benjamins.
- Jaén, M. (2007). A corpus-driven design of a test for assessing the ESL collocational competence of university students. *IJES*, 7(2), 127-147.
- Levis, J., & Cortes, V. (2008). Minimal pairs in spoken corpora: Implications for pronunciation assessment and teaching. In C. A. Chapelle, Y.-R. Chung, & J. Xu (Eds.), *Towards adaptive call: Natural language processing for diagnostic language assessment* (pp.197-208). Iowa State University.
- McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: an advanced resource book*. Routledge.
- Merino, M. (2000). *The use of corpora-based input in language testing: developing an academic vocabulary test from 'The Academic Word List' and 'The British National Corpus'*. Unpublished MA Dissertation, Lancaster University.
- Obunsha. (2002, 2003, 2005, 2011, 2012, 2021). *Zenkoku daigaku nyushi mondai seikai eigo—shiritsu dai hen*. [Compilation of English past exam papers – private university exam version]. Obunsha.
- Park, K. (2014). Corpora and language assessment: The state of the art. *Language Assessment Quarterly*, 11(1), 27-44.
- Read, J. (2002). The use of interactive input in EAP listening assessment. *Journal of English for Academic Purposes*, 1, 105-119.
- Rees, J. (1998). A new approach to rational deletion in cloze tests. *Paper read at the University of Birmingham* in 1998.
- Sharpling, G. P. (2010). When bawe meets welt: The use of a corpus of student writing to develop items for a proficiency test in grammar and English usage. *Journal of Writing Research*, 2, 175-189.
- Shillaw, J. (1994). Using a corpus to develop vocabulary tests. In L. Flowerdew & A.K.K. Tong (Eds.), *Entering Text* (pp.166-177). The Hong Kong University of Science and Technology.
- Shortall, T. (2007). The L2 Syllabus: corpus or contrivance? *Corpora*, 2(2), 157-185.
- Usami, H. (2005). *Using corpora for language testing—examining the English used in university entrance examinations in Japan with the British National Corpus and the Longman Learner Corpus*. Unpublished MA Dissertation, Lancaster University.
- Usami, H. (2012). *The application of corpora to language testing—identifying and improving problematic multiple choice grammar questions in Japanese university entrance exams*. Unpublished PhD Thesis, Lancaster University.
- Usami, H. (2015). *The application of corpus linguistics to language testing—improving multiple choice questions*. Saarbrücken, LAP Lambert Academic Publishing.
- Usami, H. (2018). A validation study of the CEFR vocabulary levels of Japanese English learners in the English Vocabulary Profile. *Tokai University: The Bulletin of the International Education Center*, 38, 35-49.
- Usami, H. (2021). Applying specialised corpora to language testing: Development of the Exam Corpus (vocabulary and grammar sections). *Tokai University: The Bulletin of the International Education Center*, 3(1), 1-13.
- Walsh, S. (2010). What features of spoken and written corpora can be exploited in creating language teaching materials and syllabuses? In A. O'Keeffe & M. McCarthy (Eds.), *The Routledge handbook of corpus linguistics* (pp. 333-344). Routledge.
- Widdowson, H. G. (2000). On the limitations of linguistics applied. *Applied Linguistics*, 21, 3-25.

要旨

一般コーパスにおける真正性と頻度を使用した語彙 4 択問題における句動詞とその問題文の調査

宇佐美 裕子

コーパスとは、言語分析のために特定の言語の代表となるよう厳選して収集された書き言葉と話し言葉の電子データである。コーパスは長い間言語教育に応用されており、特に言語テスト・評価に応用されている。コーパスの言語テスト・評価の分野における応用は 1990 年代半ばから理論的に提唱され始め、それ以来、一般コーパス、学習者コーパス、特殊コーパスの言語テスト・評価への応用に関する具体的な研究は、様々な研究者により提案され発表されてきた。さらに、このような様々な種類のコーパスは多種多様な方法で語彙テストの研究にも応用されてきている。しかし、実際に出题された過去の試験問題を収集した特殊コーパスはほとんど構築されておらず、そのような特殊コーパスは頻繁にテストされている語彙を調査するのに応用されていなければ、試験で使用される英語を調査するのに十分に応用されていない。従って、本論文は、著者によって構築された特殊コーパスである **Exam Corpus** がどのように言語テスト・評価に応用できるかを議論することを目的とする。特に、様々なコーパスを使用して、英語能力試験の語彙 4 択問題の、特に句動詞に使用されている英語を、1) 頻繁にテストされている句動詞項目、2) 頻繁にテストされている句動詞項目に提示されている問題文、の観点から調査することを目的とする。結果、句動詞の中で **put up with** が最も頻繁にテストされているが、一般コーパスに見られる実際に使用されている英語では、それほど頻繁に使用されていないことが判明した。さらに、**put up with** の問題文に使用されている英語について、共通点はあったものの、一般コーパスに見られる実際に使用されている英語と、英語能力試験で提示されている英語に幾つかの相違があったことも判明した。