

Examining the authenticity of the English used in short conversation questions in the Exam Corpus: A case study of request interrogatives using the word *mind*

Hiroko USAMI*

1. Introduction

Corpora, electronically collected written or spoken texts representing a particular language or genre for linguistic analysis (e.g. Baker et al., 2006; McEnery et al., 2006), have been utilised to describe vocabulary and grammar. These corpus-based descriptions of vocabulary and grammar have been employed in compiling teaching materials, dictionaries, wordlists, and grammars (e.g. Hunston, 2002; Chambers, 2010; Cheng, 2010; Walsh, 2010), employing authenticity, frequency, collocation, phraseology, and variation. Additionally, these corpus-based descriptions have been applied to language testing, involving the use of specialised corpora containing specific text types to inform and validate existing examinations such as TOEFL (e.g. Biber et al., 2001; 2004; Biber, 2006) and university entrance examinations in Japan (XXX, 2012; 2015).

In general, studies on corpus-based descriptions have revealed significant differences between the frequency of specific lexis and grammatical constructions in corpora compared to teaching materials. Arguments both for and against this application exist. Most researchers agree that authentic English from general corpora should be incorporated into language teaching (e.g. Hunston & Francis, 1999), whereas some corpus linguists disagree with creating materials solely based on native-speaker usage frequency.

The analyses of authenticity and frequency information derived from general corpora have been extensively discussed concerning teaching materials but not adequately addressed in language testing, particularly in English presented in examinations. Furthermore, authenticity in conversation questions from actual past English proficiency examinations has not been thoroughly examined or analysed yet, despite the application of authentic texts from general corpora being crucial to enhance construct validity in language testing.

Therefore, this study aims to inform and examine the authenticity of English frequently presented and tested in English proficiency examinations. This paper investigates the frequent presentation of multiword units and collocates as n-grams in short conversation questions from English proficiency examinations in the Exam Corpus, with a specific focus on request interrogatives using the word *mind*. The findings reveal significant differences in collocation use in *do/would you mind* between English proficiency examinations and actual authentic English.

2. Literature Review

2 – 1 Corpus-based Vocabulary and Grammar Descriptions and the Applications to Language Testing

投稿日 2024 年 11 月 20 日 受理日 2025 年 3 月 25 日

* Undergraduate School of Humanities and Culture, Tokai University

Corpora, electronically collected written or spoken texts representing a particular language or genre for linguistic analysis (e.g. Baker et al., 2006; McEnery et al., 2006), have been extensively employed to describe vocabulary and grammar. Regarding vocabulary, exploration encompasses word classes (e.g. Altenberg, 1990), individual words (e.g. *idea*, *concept*, and *notion* in Butler, 2008), collocations (e.g. Altenberg, 1991), and morphology (e.g. the *-ly* suffix in adverbs in Opdahl, 1991). For grammar, the scope includes prepositions (e.g. Kennedy, 1991), relative clauses (e.g. Lehmann, 2002), apposition (e.g. Meyer, 1991), conditionality (e.g. Wang, 1991), negation (e.g. Tottie, 1991), *get*-passives (e.g. Carter & McCarthy, 1999), and *if*-conditionals (e.g. Shortall, 2007).

The fundamental application of corpus-based descriptions lies in compiling teaching materials, dictionaries, wordlists, and grammars (e.g. Hunston, 2002; Chambers, 2010; Cheng, 2010; Walsh, 2010), leveraging additional information on authenticity, frequency, collocation, phraseology, and variation (Hunston, 2002; McEnery et al., 2006). The Longman Grammar of Spoken and Written English (Biber et al., 1999), informed by corpus data, presents grammatical features based on their frequency across four registers: fiction, academic prose, news, and conversation.

Moreover, these corpus-based descriptions extend to language testing, using specialised corpora containing specific text types, such as field-specific (e.g. business, law, aviation, and accountancy) and age-specific corpora, to inform and validate existing examinations. The TOEFL 2000 Spoken and Written Academic Language Corpus (T2K-SWAL) was instrumental in investigating university-level language skills, providing an empirically grounded alternative to the intuition of TOEFL test constructors and item writers (Biber et al., 2001; 2004; Biber, 2006). The University Entrance Examinations in Japan (UEEJ) Corpus (the precursor to the Exam Corpus), containing multiple-choice vocabulary and grammar questions from Japanese university entrance examinations, underwent analysis to examine the English presented in those examinations. Collocations in the stems and frequently targeted grammar, answers, and distractors were scrutinised alongside Japanese learners' productions in the Longman Learners' Corpus and authentic English in the British National Corpus (BNC) (XXX, 2012; 2015).

2 – 2 Authenticity of the English in Language Testing

Corpus-based descriptions have been employed in the compilation of teaching materials. Generally, such studies reveal significant differences in the frequency of particular lexis and grammatical constructions between corpora and teaching materials, sparking debates both in favour of and against this application. Most researchers concur that authentic English from general corpora should be incorporated into language teaching (e.g. Hunston & Francis, 1999). However, some corpus linguists challenge this view, arguing that authentic English is not always imperative for language teaching. Creating materials solely based on the prevalence in native-speaker usage (e.g. Widdowson, 1990; Cook, 1998; Widdowson, 2000) is considered questionable. Additionally, factors such as teachability may outweigh the importance of authentic English (e.g. Barlow, 1996; Cook, 1998; Widdowson, 2000). Widdowson (1998) contends that authentic language may not be essential in classrooms, proposing that factors beyond frequency should be considered in material compilation. Barlow (1996) and Shortall (1999) focus on 'prototypical' usages, deemed salient but not necessarily the most frequent, distinguishing these from 'typical' usages. Hunston (2002) argues that psychologically prototypical examples should be considered, advocating for the inclusion of even non-typical usages in teaching materials. McCarthy and Carter (2001) assert that the concepts of teachability and learnability are also crucial.

3. Research Questions

As discussed in the above literature, the analyses of authenticity and frequency information have been explored in the context of teaching materials but not extensively in language testing, particularly regarding the English presented in examinations. In language testing, construct validity, alongside reliability, has been acknowledged as crucial. Consequently, there is a need for the increased application of authentic texts from general corpora in language testing to enhance construct validity. In English proficiency examinations, aside from those categorised across proficiency levels, the repetition of targeted grammatical categories, vocabulary, or specific stems may lead test-takers to perceive them as significant, assuming that the frequency of presentation reflects importance in actual English. Moreover, the authenticity employed in conversation questions from past English proficiency examinations has not undergone sufficient examination or analysis.

Therefore, this study aims to inform and scrutinise the authenticity of English frequently presented and tested in English proficiency examinations. This paper investigates the frequent presentation of multiword units and collocates in short conversation questions from English proficiency examinations in the Exam Corpus, with a specific focus on request interrogatives using the word *mind*. The following research questions (RQs) were posed:

RQ1: What multiword units and collocates are frequently used for request interrogative using the word *mind* in the short conversation questions in the Exam Corpus?

RQ2: What collocates of the word *mind* are frequently used in authentic English conversations in a spoken corpus?

4. Method

4 – 1 The Exam Corpus

In this study, the Exam Corpus was analysed to explore multiword units and collocates in English proficiency examinations and address RQ1. The Exam Corpus is a specialised corpus constructed by the author for application in language testing and the examination of existing English proficiency examinations. Table 1 displays the number of questions across each skill and examination in the Exam Corpus (Ver. 1.5).

Table 1

Number of question items across each skill and examination in the Exam Corpus (Ver. 1.5)

Skill / Exam	UEEJ	TEAP	EIKEN	GTEC	TOEFL	TOEIC	CMB
Reading	✓	✓	✓	✓	✓	✓	✓
Listening	✓	✓	✓	✓	✓	✓	✓
Long conversation	2,436	-	4	-	✓	✓	✓
Short conversation	3,193	-	190	-	-	-	-
Grammar	11,545	0	426	0	176	1,156	0
Vocabulary	10,075	60	2,099	92	0	604	380

Note. UEEJ = University Entrance Examinations in Japan. TEAP = Test of English for Academic Purposes. GTEC = Global Test of English Communication. TOEFL = Test of English as a Foreign Language. TOEIC = Test of English for International Communication. CMB = Cambridge English.

The Exam Corpus (Ver. 1.5) contains 745,692 words presented in 27,249 exclusively multiple-choice vocabulary, grammar, and short and long conversation questions from four kinds of English proficiency

examinations, taken from only the official test books published after 2000 by authorised examination boards or publishers: 1) University Entrance Examinations in Japan (UEEJ) for 2001–2024, 2) English proficiency examinations administered in Japan (Test of English for Academic Purposes (TEAP) for 2015–2016, EIKEN—pre-1st, 2nd, pre-2nd and 3rd grades for 2011–2024, and Global Test of English Communication (GTEC) for 2018–2019), 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, and Test of English for International Communication (TOEIC) for 2005–2023); and 4) English proficiency examinations created by Cambridge Assessment English in the U.K. (Cambridge English–Key English Test (KET), Preliminary English Test (PET) and First Certificate in English (FCE) for 2006–2014). The Exam Corpus (Ver. 1.5) currently contains only vocabulary, grammar, and short and long conversations questions but will add question items presented in reading and listening set in the near future, as indicated as the symbol  in the Table 1.

4 – 2 Short Conversation Questions in The Exam Corpus

In the Exam Corpus, short conversation questions were categorised as ones, including up to four turns, whereas long conversation questions included more than five turns. Presenting conversations up to four turns without gap-filling questions or those with gap-filling questions did not matter. In this study, a focused analysis was conducted on 157 short conversation questions (2 questions from EIKEN Pre-2nd and 155 questions from UEEJ) that included the word *mind*, whether it is a verb or a noun, (token is 7,107, type is 1,150, and TTR is 16.1) from UEEJ and EIKEN Pre-2nd and 3rd grades within the Exam Corpus, encompassing all the skills and English proficiency examinations present.

4 – 3 #LancsBox

The software package #LancsBox (see <http://corpora.lancs.ac.uk/lancsbox/>) developed at Lancaster University was used in order to analyse the short conversation questions in the Exam Corpus. Among the functions available in #LancsBox, the functions called *Ngrams* (for n-grams) and *KWIC* (for concordances) were used in this study.

4 – 3 The Spoken BNC2014

To investigate authentic English and address RQ2, the Spoken BNC2014, a spoken component of the BNC2014, containing 11.5 million words from 1,251 informal spoken interactions by 672 speakers of British English collected between 2012 and 2016 (Love et al. 2017) the data was utilised and analysed using #LancsBox.

5. Analysis and Discussion

5 – 1 N-grams of *mind* in the Short Conversation Questions

To explore the multiword units employed with request interrogatives using *mind* in the short conversation questions within the Exam Corpus, 4-, 5-, and 6-grams were extracted. As analysed in XXX (2023), these multiword units are incorporated as a part of interrogative sentences (e.g. *you mind if I, do you mind if, do you mind if I, I would you mind if I*) and ones presented in a part of the interrogative (e.g. *A do you mind if, A do you mind if I, A would you mind if I*). This section delves deeper into the presentation of *mind* in the short conversation questions within the Exam Corpus, specifically obtaining 4-, 5-, and 6-grams solely including *mind* from this subset of questions.

Examining the authenticity of the English used in short conversation questions in the Exam Corpus:

A case study of request interrogatives using the word *mind*

Table 2

4-, 5-, and 6-grams of the short conversation questions including *mind* in the Exam Corpus

Rank	4-gram	Freq.	5-gram	Freq.	6-gram	Freq.
1	you mind if I	34	do you mind if I	26	a do you mind if I	11
2	do you mind if	28	a do you mind if	11	a would you mind if I	6
3	a do you mind	17	would you mind if I	8	you mind if I sit here	4
4	a would you mind	16	a would you mind if	6	do you mind if I smoke	4
5	would you mind if	8	mind a do you mind	5	mind a do you mind if	4
6	if you don't mind	7	mind if I sit here	4	do you mind if I sit	3
7	no not at all	7	you mind if I sit	4	mind if I open the window	3
8	the window b zzz	7	you mind if I smoke	4	do you mind if I use	3
9	mind a do you	5	don't mind if I do	3	own business a do you mind	3
10	you change your mind	5	would you mind opening the	3	mind if I sit here b	3
11	mind your own business	4	no not at all no	3	woman do you mind if I	3
12	mind if I smoke	4	you mind if I ask	3	would you mind opening the window	3
13	if I sit here	4	I sit here b zzz	3	you mind if I open the	3
14	mind if I sit	4	you mind if I open	3	do you mind if I open	3
15	do you want to	4	business a do you mind	3	if I sit here b zzz	3
16	are you going to	4	you mind if I use	3	if you change your mind a	2
17	would you mind opening	3	own business a do you	3	a would you mind opening the	2
18	a would you like	3	mind if I open the	3	a zzz b I don't mind	2
19	b zzz yes I	3	if I open the window	3	but do you mind if I	2
20	mind if I do	3	if I sit here b	3	your own business a do you	2
21	would you mind taking	3	you mind opening the window	3	do you mind if I ask	2
22	you mind opening the	3	opening the window b zzz	3	you mind if I smoke here	2
23	make up my mind	3	woman do you mind if	3	do you mind if I turn	2
24	I don't mind the	3	rather you did n't a	2	I don't mind if I do	2
25	mind if I open	3	I don't think so a	2	you mind if I use your	2
26	opening the window b	3	are you going to do	2	I can't make up my mind	2
27	I help you b	3	are you sure you don't	2	mind if I smoke here b	2
28	not at all I	3	I've made up my mind	2	mind your own business a do	2
29	a I'd like to	3	mind your own business a	2	mind opening the window b zzz	2
30	mind if I use	3	no not at all yes	2	but I have to apply for	2
31	rather you did n't	3	it completely slipped my mind	2	I have to apply for it	2
32	not at all no	3	b zzz a never mind	2	in case you change your mind	2
33	business a do you	3	do you mind if we	2	b zzz a not at all	2
34	do you think I	3	your own business a do	2	business a do you mind if	2
35	sit here b zzz	3	you mind if I turn	2	I've made up my mind to	2
36	what do you think	3	to apply for it first	2	don't mind if I do a	2
37	mind opening the window	3	can't make up my mind	2	b zzz no not at all	2
38	may I help you	3	zzz b I don't mind	2	you mind opening the window b	2
39	no go ahead yes	3	b zzz not at all	2	no I don't think I can	2
40	mind if I ask	3	zzz if you don't mind	2	have to apply for it first	2

As highlighted in XXX (2023), the multiword units *do you mind if I* and *would you mind if I* were frequently observed in Table 2, with *do you mind if* occurring 28 times and *do you mind if I* appearing 26 times, significantly more than their *would you mind if* and *would you mind if I* counterparts, each occurring 8 times. Consequently, the modal *do* exhibited a much more frequent collocation with *you mind if I* than the modal *would* in the short conversation questions.

Examining the verbs that follow *do you mind if I*, various options are presented; *ask, borrow, call, change, close, eat, get, give, join, open, sit, smoke, turn, and use*. Among them, the multiword units *do you mind if I open the window?* (3 times), *do you mind if I sit here?* (3 times), *do you mind if I smoke* (4 times), and *do you mind if I use* (3 times) are frequently presented, as shown in Figure 1.

KWIC	GraphColl	Whelk	Words	Ngrams	Text	Wizard
Corpora KWIC: do you mind...						
Search						
Search do you mind if I Occurrences 26 (36.58) Texts 1 Corpus Corpus 1 Context 7 Display Text						
Index	File	Left	Node	Right		
12	2014 Kiyo d.	up? Could you play something? A: Bob,	do you mind if I	ask Mes. Brown to come over for		
19	2014 Kiyo d.	on how people spend their free time.	Do you mind if I	ask you a few questions? Susan: Oh,		
16	2014 Kiyo d.	at all. Where are you from? A:	Do you mind if I	borrow your book? B: Zzz Can you		
23	2014 Kiyo d.	I'm late. The train was delayed. A:	Do you mind if I	call you the first thing Monday morning?		
11	2014 Kiyo d.	your mind mind your own business A:	Do you mind if I	change the channel? B: Zzz The chahhel		
18	2014 Kiyo d.	near future. No, I don't mind. A:	Do you mind if I	close the window? B: Zzz Please go		
4	2014 Kiyo d.	me alone Can't you ask someone else	Do you mind if I	don't? "Can you ask me later A:		
26	2014 Kiyo d.	that. It's difficult to change my mind.	Do you mind if I	eat them all? "What do you have		
22	2014 Kiyo d.	I tell you more about a suit?	Do you mind if I	get dressed? "May I try it on?		
17	2014 Kiyo d.	moment. "If you change your mind A:	Do you mind if I	give you a word of advice? B:		
21	2014 Kiyo d.	nervous. She could change her mind. A:	Do you mind if I	join you? B: Zzz A: I'm Ted.		
24	2014 Kiyo d.	so upset Mind your own business A:	Do you mind if I	open the window? B: Zzz No, thank		
13	2014 Kiyo d.	I'll have the steak and lobster. Mary:	Do you mind if I	open the window? Bill: Zzz, please go		
2	2014 Kiyo d.	back. "Not in the same way Woman:	Do you mind if I	open the window? Man: Not a bad		
3	2014 Kiyo d.	you say? How do you feel? A:	Do you mind if I	sit here? B: Zzz Oh, do you?		
20	2014 Kiyo d.	suppose it's okay. A: Excuse me, but	do you mind if I	sit here? B: Zzz The seat is		
1	2014 Kiyo d.	"Are you all right my sitting here?	Do you mind if I	sit here? May I have seated here?		
15	2014 Kiyo d.	I'd love to. Yes, please sit. A:	Do you mind if I	smoke here? B: I'd rather you did		
9	2014 Kiyo d.	when you are finished. A: Excuse me.	Do you mind if I	smoke here? B: Zzz Yes, go ahead.		
6	2014 Kiyo d.	have any matches? Do you smoke cigarettes?	"Do you mind if I	smoke? A: I'd like to take Mr.		
14	2014 Kiyo d.	don't mind. Was it an accident? A:	Do you mind if I	smoke? B: Actually, zzz A: No problem.		
10	2014 Kiyo d.	I. Sally: Oh, you're reading here, but	do you mind if I	turn on the TV? There is a		
8	2014 Kiyo d.	you to invite me Never mind Woman:	Do you mind if I	turn the television on? Man: I am		
25	2014 Kiyo d.	you fill out this form? B: Yes,	Do you mind if I	use this pen? A: Zzz B: Thanks		
7	2014 Kiyo d.	welcome. "Not at all. Never mind. A:	Do you mind if I	use your bicycle over the weekend? B:		
5	2014 Kiyo d.	you? Attendant: Can I help you? Woman:	Do you mind if I	use your restroom? Attendant: Sure. This is		

Figure 1

Concordance of *do you mind if I* in the Exam Corpus

Conversely, Table 2 reveals that the multiword unit *would you mind* collocates with an *if*-clause (*would you mind if, would you mind if I, A would you mind if, and A would you mind if I*). However, the concordance displayed in Figure 2 suggests that *would you mind* appears more frequently with a gerund rather than an *if*-clause: *would you mind if I* occurs 8 times, whereas *would you mind* + gerund occurs 22 times. As Figure 2, *would you mind* frequently collocates with actions expressed as gerunds, such as *opening the window* (3 times).

KWIC	GraphColl	Whelk	Words	Ngrams	Text	Wizard
Corpora KWIC: would you mind						
Search						
Search would you mind Occurrences 30 (42.21) Texts 1 Corpus Corpus 1 Context 7 Display Text						
Index	File	Left	Node	Right		
3	2014 Kiyo d.	get some fresh air in here. A:	Would you mind	checking your ticket? I think you're in		
22	2014 Kiyo d.	sir, you'll have to buy something. A:	Would you mind	getting some things for me at the		
5	2014 Kiyo d.	There we are There you are A:	Would you mind	having this package sent to this address?		
29	2014 Kiyo d.	won't mind if you just listen. A:	Would you mind	helping me move my things to my		
13	2014 Kiyo d.	Don't you think so? "No problem. A:	Would you mind	if I ask your price range? B:		
16	2014 Kiyo d.	loose. "It just slipped my mind. A:	Would you mind	if I closed the window? B: Zzz		
27	2014 Kiyo d.	have to apply for it first. A:	Would you mind	if I cut in line? I'm in		
18	2014 Kiyo d.	far. You'd better take a train. A:	Would you mind	if I opened the window? B: Zzz		
14	2014 Kiyo d.	I can try. Did you quit playing?	Would you mind	if I play? Will you ever give		
17	2014 Kiyo d.	in mind? A: Hi. I'm new here.	Would you mind	if I sat at this table? B:		
9	2014 Kiyo d.	"would you do me a favor A:	Would you mind	if I sit here? B: Zzz I		
24	2014 Kiyo d.	dressed? "May I try it on? A:	Would you mind	if I used your computer to check		
12	2014 Kiyo d.	Would you mind taking some aspirin? Andy:	Would you mind	lending me your car tonight? Yutaka: Zzz		
30	2014 Kiyo d.	don't mind. It's up to you. A:	Would you mind	looking at my plan for out trip?		
14	2014 Kiyo d.	still buy tickets for tonight's show? A:	Would you mind	not playing the radio so loud? B:		
1	2014 Kiyo d.	Wouldn't it be nice seating here? A:	Would you mind	opening the window? B: Zzz "Not at		
25	2014 Kiyo d.	down? I hope you don't mind. A:	Would you mind	opening the window? B: Zzz This is		
15	2014 Kiyo d.	I'm very thin. "Yes, I am. Shall	Would you mind	opening the window? Yes, I am. Shall		
19	2014 Kiyo d.	I mind. Go ahead. Member of staff:	Would you mind	putting out your cigarette? This is a		
28	2014 Kiyo d.	would you buy another present for him?	would you mind	sending another present to him now? A:		
20	2014 Kiyo d.	it on over there if you like.	"Would you mind	showing me that blue dress? Please show		
10	2014 Kiyo d.	any goods to declare? B: Zzz A:	Would you mind	showing them to me, please? I'm a		
23	2014 Kiyo d.	the 9:45 flight? Airline company: Zzz Sure.	Would you mind	sitting down on the flight? "Yes. There		
21	2014 Kiyo d.	do. How do you go there? Harry:	Would you mind	taking a look at this? Nick: This		
26	2014 Kiyo d.	often? Are you still looking for it?	"Would you mind	taking my seat instead? Bob: What happened?		
11	2014 Kiyo d.	aspirin? "Why don't you take some aspirin?	Would you mind	taking some aspirin? Andy: Would you mind		
6	2014 Kiyo d.	I should take it next year instead?	would you mind	talking to me in another room? A:		
2	2014 Kiyo d.	Yes, it's too hot. A: Excuse me.	Would you mind	trading seats with me? B: Zzz. I		
7	2014 Kiyo d.	for you if only I knew A:	Would you mind	watching my suitcase while I go to		
4	2014 Kiyo d.	"No problem. Let's take a break. Naomi:	Would you mind	zzz tomorrow? My car is being repaired.		

Figure 2

Concordance of *would you mind* in the Exam Corpus

A case study of request interrogatives using the word *mind*

Furthermore, instances of the multiword unit with negation (*if you don't mind, don't mind if I do, zzz if you don't mind, a zzz b I don't mind, I don't mind if I do, don't mind if I do a*) were identified.

Moreover, alongside request interrogatives with *mind*, the multiword unit *not at all*, seemingly employed as a response to request interrogatives, is also presented (*no not at all, not at all I, not at all no, no not at all no, no not at all yes, b zzz not at all, b zzz a not at all, b zzz no not at all*), indicated in Table 2.

5 – 2 Collocates of *mind* in the Spoken BNC2014 and the Exam Corpus

To further scrutinise how the word *mind* is utilised in request interrogatives, the collocates of *do you mind* and *would you mind* were extracted from the Spoken BNC2014 and the short conversation questions in the Exam Corpus.

Table 3

The collocates of *do you mind* and *would you mind* in the Spoken BNC2014 and the Exam Corpus

The Spoken BNC2014				The Exam Corpus			
Do you mind	149	Would you mind	41	Do you mind	39	Would you mind	30
Do you mind -ing	34	Would you mind -ing	17	Do you mind -ing	5	Would you mind -ing	21
Do you mind me -ing	7	Would you mind me -ing	0	Do you mind my -ing	3	Would you mind my -ing	0
Do you mind if	75	Would you mind if	15	Do you mind if	28	Would you mind if	8
Do you mind if I	61	Would you mind if I	9	Do you mind if I	26	Would you mind if I	8

As shown in Table 3, *do you mind* is notably more frequent than *would you mind* in the Spoken BNC2014, whereas the frequency difference between the two is less pronounced in the Exam Corpus.

As for the words that follow *do you mind* and *would you mind*, an *if*-clause is more frequently observed than a gerund in both the Spoken BNC2014 and the Exam Corpus. For *would you mind*, however, the frequency difference between gerunds and *if*-clauses is less marked in the Spoken BNC2014, but gerunds are significantly more frequent than *if*-clauses in the Exam Corpus.

Regarding gerunds' subject, *me* is more commonly used in the Spoken BNC2014, whereas *my* tends to be more frequent in the Exam Corpus, although the frequency of gerunds' subject is relatively low. Notably, there are no instances of *would you mind* + any gerunds' subject in both the Spoken BNC2014 and the Exam Corpus.

Concerning *if*-clauses, they tend to collocate more frequently with *do you mind* than with *would you mind* in both the Spoken BNC2014 and the Exam Corpus. Additionally, the subject *I* is commonly used in both corpora.

Next, to delve into the verbs that collocate with *mind* in request interrogatives, the collocate verbs in *do you mind* and *would you mind* were extracted from the Spoken BNC2014 and the short conversation questions in the Exam Corpus.

Table 4

Collocate verbs in *do you mind* in the Spoken BNC2014 and the Exam Corpus

The Spoken BNC2014				The Exam Corpus			
<i>if</i> clause		gerund		<i>if</i> clause		gerund	
do	6	going	6	smoke	4	giving	2
have	6	asking	3	open	3	carrying	1
put	4	passing	2	sit	3	changing	1
smoke	3	putting	2	use	3	helping	1
turn	3	signing	2	turn	2	moving	1
borrow	2	taking	2	ask	2	dropping	1
pay	2					opening	1
record	2						

Table 4 illustrates the verbs that frequently follow *do you mind* in both the Spoken BNC2014 and the Exam Corpus. Notably, *smoke* and *turn* are common collocates in both corpora. However, substantial differences exist in the range of verbs associated with *do you mind* between the two. In the Spoken BNC2014, for example, the verbs *do*, *have*, and *put* frequently appear in *if*clauses, whereas these verbs are absent in the Exam Corpus. Additionally, gerunds such as *going* and *asking* are prevalent in the Spoken BNC2014, but do not occur in the Exam Corpus, which displays a diverse array of gerunds. Conversely, *open*, *take*, and *show* frequently collocate with *do you mind* in the Exam Corpus but are not present in the Spoken BNC2014.

Table 5

Collocate verbs in *would you mind* in the Spoken BNC2014 and the Exam Corpus

The Spoken BNC2014				The Exam Corpus			
<i>if</i> clause		gerund		<i>if</i> clause		gerund	
have	2	putting	2	ask	1	opening	3
wait	2	doing	2	closed	1	taking	3
went	2	being	1	cut	1	showing	2
do	1	flying	1	opened	1	checking	1
boil	1	giving	1	play	1	getting	1
keep	1	going	1	sat	1	having	1
lay	1	having	1	sit	1	helping	1
parked	1	following	1	used	1	lending	1
stick	1	getting	1			looking	1
throw	1	putting	1			playing	1
borrowed	1	keeping	1			putting	1
chat	1	picking	1			sending	1
		popping	1			sitting	1
		reading	1			talking	1
		writing	1			trading	1
		standing	1			watching	1
		taking	1				

Table 5 displays the collocates of *would you mind* in the Spoken BNC2014 and the Exam Corpus. The verbs *have*, *wait*, and *went* are commonly associated with *would you mind* in *if*clauses in the Spoken BNC2014. Conversely, in the Exam Corpus, gerunds such as *opening*, *taking*, and *showing* frequently follow *would you mind*.

6. Conclusion

This study has scrutinised and deliberated upon the authenticity of the English presented in short conversation questions within the Exam Corpus, with a specific emphasis on how the request interrogative using *mind* is articulated in comparison to the authentic English found in a general spoken corpus.

First, echoing Usami (2023), the frequent presentation of multiword units such as *do you mind if I* and *would you mind if I* was evident. However, *do you mind if* and *do you mind if I* were considerably more prevalent than their counterparts *would you mind if* and *would you mind if I*. An examination of collocations following *do you mind if I* revealed a variety of verbs, including *ask*, *borrow*, *call*, *change*, *close*, *get*, *give*, *join*, *open*, *sit*, *smoke*, *turn*, and *use*. Notably, the specific multiword units *do you mind if I open the window*, *do you mind if I sit here*, *do you mind if I smoke*, and *do you mind if I use* were frequently encountered. Conversely, the concordance analysis indicated a higher frequency of *would you mind* with gerunds rather than *if*clauses, often collocating with *opening the window*. Additionally, multiword units incorporating gerund subjects (*do you mind my*), negation (*if you don't mind*), and the multiword unit *not at all* as a response to request interrogatives were also identified.

Second, a detailed exploration of how the word *mind* is presented in request interrogatives revealed a marked discrepancy between *do you mind* and *would you mind* in the Spoken BNC2014, with the former being significantly more frequent. The Exam Corpus displayed a lesser contrast in frequency between the two expressions. Moreover, *if*clauses were observed to more frequently collocate with *do you mind* in both corpora, whereas there was less disparity in frequency between gerunds and *if*clauses with *would you mind* in the Spoken BNC2014, compared to a more pronounced preference for gerunds over *if*clauses in the Exam Corpus. Furthermore, the gerund subjects *me* and *my* exhibited varying frequencies, with *me* being more frequent in the Spoken BNC2014 and *my* being more prevalent in the Exam Corpus, despite both having a relatively low frequency. *If*clauses were more inclined to collocate with *do you mind* than *would you mind* in both corpora, with the subject *I* being frequently associated with both expressions. Examining collocated verbs, *smoke* and *turn* were identified after *do you mind if* in both corpora. However, the verbs *do*, *have*, and *put* were prominent in *do you mind + if* clause, and the gerunds *going* and *asking* were frequent in the Spoken BNC2014. In contrast, a diverse array of gerunds, including *having*, *getting*, and *putting*, was observed with *would you mind*. Notably, the verbs *have*, *wait*, and *went* were frequently associated with *if*clauses in the Spoken BNC2014, whereas the gerunds *opening*, *taking*, and *showing* were prevalent in the Exam Corpus.

In conclusion, this study analysed short conversation questions extracted from university entrance examinations and EIKEN in Japan within the Exam Corpus. However, future investigations could profitably extend beyond short conversation questions to encompass long conversation questions, specifically those with more than five turns, alongside an exploration of their n-grams, concordance, and keywords. This broader examination, spanning both general corpus conversations and English proficiency examination contexts, promises a more comprehensive understanding of authentic English usage.

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.

References

- Altenberg, B. (1990). Spoken English and the dictionary. In Svartvik, J. (Ed.) *The London-Lund Corpus of Spoken English: Description and Research*, pp. 177-192. Lund University Press.
- Altenberg, B. (1991). A bibliography of publications relating to English computer corpora. In Johansson, S. & A.-B. Stenström (Eds.) *English Computer Corpora: Selected Papers and Research Guide*, pp. 355-396. Mouton de Gruyter.
- Baker, P., Hardie, A., & McEnery, T. (2006). *A glossary of corpus linguistics*. Edinburgh University Press.
- Barlow, M. (1996). Corpora for theory and practice. *International Journal of Corpus Linguistics* 1: 1-37.
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. Longman.
- 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 Simpson, R. C. & 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.
- Biber, D. (2006). *University language: A corpus-based study of spoken and written registers*. John Benjamins.
- Butler, C. S. (2008). The very idea! A corpus-based comparison of IDEA, CONCEPT and NOTION and their formal equivalents in Spanish. *Journal of the Spanish Association of Anglo-American Studies* 30(2): 59-77.
- Carter, R. & McCarthy, M. (1999). The English *get*-passive in spoken discourse: description and implication for an interpersonal grammar. *English Language and Literature* 3/1: 41-58.
- Chambers, S. (2010). What is data-driven learning? In O'Keeffe, A. & 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 O'Keeffe, A. & M. McCarthy (Eds.) *The Routledge Handbook of Corpus Linguistics*, pp. 319-332. Routledge.
- Cook, G. (1998). The uses of reality: a reply to Ronald Carter. *ELT Journal* 52(1): 57-64.
- 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*. John Benjamins.
- Kennedy, G. (1991). *Between and through*: the company they keep and the functions they serve. In Aijmer, K. & B. Altenberg (Eds.) *English Corpus Linguistics: Studies in Honour of Jan Svartvik*, pp. 95-110. Longman.
- Lehmann, H. (2002). Zero subject relative constructions in American and British English. In Peters, P., Collins, P. & A. Smith (Eds.) *New Frontiers in Corpus Research*, pp. 163-177. Rodopi.
- Love, R., Dembry, C., Hardie, A., Brezina, V., & McEnery, T. (2017). The Spoken BNC2014. *International Journal of Corpus Linguistics* 22(3): 319-344.
- McCarthy, M. & Carter, R. (2001). Size isn't everything: spoken English, corpus, and the classroom. *TESOL Quarterly* 35(2): 337-340.
- McEnery, T., Xiao, R., & Tono, Y. (2006). *Corpus-based language studies: an advanced resource book*. Routledge.
- Meyer, C.F. (1991). *Apposition in Contemporary English*. Cambridge University Press.
- Opdahl, L. (1991). -ly as adverbial suffix: corpus and elicited material compared. *ICAME Journal* 15: 19-35.
- Shortall, T. (1999). Protogrammar, frequency, and the acquisition of structure. *Workshop given at the 25th JALT conference*, Maebashi, Japan.
- Shortall, T. (2007). The L2 Syllabus: corpus or contrivance? *Corpora* 2(2): 157-185.

Examining the authenticity of the English used in short conversation questions in the Exam Corpus:

A case study of request interrogatives using the word *mind*

Tottie, G. (1991). *Negation in English Speech and Writing: A Study in Variation*. Academic Press.

XXX. (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.

XXX. (2015). *The application of corpus linguistics to language testing—improving multiple choice questions*. Saarbrücken, LAP Lambert Academic Publishing.

Walsh, S. (2010). What features of spoken and written corpora can be exploited in creating language teaching materials and syllabuses? In O'Keeffe, A. & M. McCarthy (Eds.) *The Routledge handbook of corpus linguistics*, pp. 333-344. Routledge.

Wang, S. (1991). *A Corpus Study of English Conditionals*. Unpublished MA thesis, Victoria University of Wellington.

Widdowson, H.G. (1990). *Aspects of Language Teaching*. Oxford University Press.

Widdowson, H.G. (1998). Context, community, and authentic language. *TESOL Quarterly* (32)4: 705-716.

Widdowson, H.G. (2000). On the limitations of linguistics applied. *Applied Linguistics* 21(1): 3-25.

要旨

Exam Corpus の会話問題に使用される英語の真正に関する調査

XXX XXX

コーパス（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）は語彙と文法を記述するために使用されている。コーパスを元にした語彙と文法の記述は、コーパスから得られる真正、頻度、コロケーション、フレーズ、バリエーションの情報を活かして、教材、辞書、語彙リスト、文法書の編纂に使用されており、さらに特定の種類の英語を収集して構築された特殊コーパスを用いて、TOEFL や日本の大学入試問題などの英語検定試験の妥当性を証明するなど、言語テストの分野にも応用されている。概して、コーパスを元にした語彙と文法の記述は、教材に提示されているものと比較して、特定の語彙と文法の頻度に違いがあることを示している。ほとんどの研究者は一般コーパスから得られた実際に使用されている英語は英語教育に活かされるべきだと主張しているが、英語教材はネイティブスピーカーによる英語の使用頻度を元にしてだけ作成されるべきではないと主張する研究者もいる。一般コーパスから得られた真正や頻度情報の分析は、教材開発の観点からは十分に議論されているが、言語テストの分野、特に英語検定試験に提示されている英語においては十分議論されてはいない。さらに、実際の英語検定試験における会話問題に見られる真正は、十分に調査、分析されていない。従って、本論文は英語検定試験で提示され、問われている英語の真正を調査することを目的とする。本論文では、著者が構築した特殊コーパスである Exam Corpus に収集された英語検定試験の短い会話問題に出題された mind を使用した依頼をする疑問文に焦点を当て、n-gram に現れた multiword unit とコロケーションを調査した。結果、英語検定試験と一般コーパスに見られる実際に使用されている英語において、do/would you mind におけるコロケーションに違いがあることが判明した。