Contrastive Interlanguage Analysis of Modal Auxiliary Verb Usage by Japanese Learners of English in Argumentative Essays

Shusaku Nakayama, Meiji Gakuin University, Japan

The IAFOR International Conference on Education – Hawaii 2020
Official Conference Proceedings

Abstract
This research argues that in writing Japanese non-native English speakers use English modal auxiliary verbs in a way that significantly differs from how native English-speaking students/teachers use them from two different perspectives: frequency of use and verb phrase structures (VPS) where modal verbs can occur. To this hypothesis, the use of nine central modals (*can*, *could*, *may*, *might*, *shall*, *should*, *will*, *would*, and *must*) by Japanese learners was compared with that by native English-speaking students and native English-speaking teachers respectively. This comparison was carried out using the International Corpus Network of Asian Leaners of English, which is one of the largest freely-available corpora of Asian learners' English. Frequency analysis revealed Japanese college students' overuse of *can*, *should*, and *must* as well as underuse of *will* and *would* as compared to native English speakers. VPS analysis revealed that Japanese students and native English-speakers shared different preferences for VPSs. Japanese students infrequently used the modals in the progressive and perfect aspect relative to native English-speakers. Overall findings suggest that teaching materials or language teachers should explain other modality items so that learners can have a wide range of lexical items to reflect their feelings more accurately, and should not teach them in exactly the same way because depending on the modals, preferences for VPSs were different.

Keywords: modal auxiliary verbs, corpus linguistics, contrastive interlanguage analysis, Japanese learners of English, the ICNALE corpus

iafor
The International Academic Forum
www.iafor.org

Introduction

For non-native English speakers, modal auxiliary verbs (e.g., *can, will, might*) are among the most challenging grammatical structures to master, something that has been widely recognized by many researchers (Cook, 1978; Decapua, 2008; Khojasteh & Kafipour, 2012; Mukundan & Khojasteh, 2011; Römer, 2004). The difficulty in using modal verbs can be attributed to their wide range of meanings (Holmes, 1988) or their unique rules about forms (Celce-Murica & Larsen-Freeman, 1999). Briefly explained, each modal has at least two different meanings and does not have inflected forms as in *she cans play the guitar* or *she canned play the guitar*.

Studies related to leaner's modal verb usage have been conducted for a number of non-native English speakers; for example, for Swedish (Aijmer, 2002), Chinese (Yang, 2018; Xiao, 2017), Indian (Wilson, 2005), Malaysian (Khojasteh & Reinders, 2013), Brazilian (Viana, 2006) and so forth. These studies all come to largely the same conclusion that learners' modal usage is problematic, one of the reasons that motivated me to conduct this study. Despite these facts, however, there are few studies in relation to the Japanese learners' modal usage.

One of the studies focusing on Japanese learner's modal usage is conducted by Fujimoto (2019). She compared the use of epistemic modal verbs (e.g., might, could) and epistemic adverbs (e.g., perhaps, probably) employed by Japanese university students with that by American/British English speakers in written corpora. Research findings indicated that Japanese students frequently used the modal verbs while they infrequently used the adverbs. In this research, one point that we can improve could lie in the methodology. Specifically, Japanese learner's language use was analyzed in a corpus that she created from writing assignments produced by students taking her academic writing course whereas English native speaker's language data was obtained from the existing corpora, which is available at https://cqpweb.lancs.ac.uk/. Differences in writing conditions including topics or allocated time to complete writing tasks can largely affect their products (Ishikawa, 2013). It would be, therefore, worth conducting a study on Japanese learner's modal verb usage in writing whose conditions are better controlled.

To obtain a clear picture of Japanese learner's modal verb usage, this research is carried out based on the concept of *Contrastive Interlanguage Analysis* (CIA), which was firstly proposed by Granger (1996). She explains that "CIA does not establish comparisons between two different languages but between native and learner varieties of one and the same language" (p. 43). When the focus of research is on learner language, she sees two types of comparison as being worthwhile: one is a comparison between learner language and native language, and the other is a comparison of different interlanguages of the same language. This research adopts the former approach to uncover characteristics of Japanese learner language.

Literature Review

This chapter introduces what English modal verbs are, and how they have been studied in the field of corpus linguistics.

Modal Auxiliary Verbs

English modals contribute to the role of making our expressions richer or expressing our own perspectives appropriately. Modal verbs are used

> "to give a proposition a degree of probability, to express one's attitude, and to perform various social functions, such as expressing politeness or indirectness when making requests, giving advice, or granting permission. […] When English speakers use a modal, they interject their own perspective and view a proposition more subjectively than when they simply use present or past tense. (Celce-Murica & Larsen-Freeman, 1999, p. 141)"

What this statement indicates is that speakers' or writers' lack of knowledge regarding modal verbs would mean that they might always end up expressing their propositions directly even if they would like to make their opinions more indirectly.

According to researchers (Coates, 1983; Depraetere & Reed, 2006; Kennedy, 2002), modal verbs are generally divided into three kinds of sub-classes: nine central modals (*will*, *would*, *can*, *could*, *may*, *might*, *shall*, *should*, *must*); semi-modals (*want to*, *be going to*); marginal modals (*need to*, *ought to*, *used to*). The use of the nine central modals is, in fact, decreasing year by year whereas that of the semi-modals is increasing in written English as exemplified by the Brown family of corpora (Leech et al, 2009). At the same time, they argue that the overall frequency of the nine central modals is much higher than that of the semi-modals, meaning that it would be valid to study the nine central modals as a first step for studies on Japanese learners' modal usage.

Verb Phrase Structure

Kennedy (2002) maps nine verb phrase structures (hereafter, VPS) where the nine central modal verbs can occur. In this paper I will use S1-S9 in order to refer to each of the VPSs.
S1: Modal alone (e.g. Who will go? I will.)
S2: Modal + bare infinitive (e.g. She will go there.)
S3: Modal + be + past participle (e.g. It should be done.)
S4: Modal + be + present participle (e.g. I should be cleaning the room now.)
S5: Modal + have + past participle (e.g. I should have done it.)
S6: Modal + be + being + past participle/adjective (e.g. This room should be being cleaned.)
S7: Modal + have + been + past participle (e.g. This door should have been fixed.)
S8: Modal + have + been + present participle (e.g. We should have been waiting for her.)
S9: Modal + have + been + being + past participle/adjective (He might have been being careless.)
The key point in VPS analyses is to identify differences in preferences that the modal verbs have for the VPSs between Japanese learner language and native language.

Corpus-based studies on modal verbs

Corpus linguists seem to have similar interests in modal verbs. Römer (2004) compared the use of the central nine modals occurring in the real world as exemplified by the British National Corpus (BNC) with that in German textbooks from three perspectives: frequency of occurrence, modal meanings, and syntactic surroundings (negative sentences, questions, set phrases, if-clauses, and passive construction) where each modal tends to co-occur. In conclusion, she made several suggestions for the treatment of the modals in textbooks to fill in gaps between the English used in the textbooks and that in the real world.

Khojasteh and Kafipour (2012) investigated to what extent Malaysian English textbooks reflected the modal use in the real world from the point of view of VPSs, and this was carried out by comparing the modal usage in BNC with that in a textbook corpus. In addition to their corpus-based findings, they used three more corpus-based findings for the purpose of the research: Kennedy's (2002), Mindt's (1995), and Mindt's (2000). Based on these, they revealed differences in frequency of contribution of each modal to each VPS between the English used in the real world and that in Malaysian textbooks. Consequently, they considered the underuse of several VPSs in the textbooks as being problematic and concluded that Malaysian textbook developers had ignored corpus-based findings.

Nordberg (2010) investigated if the use of the central nine modals occurring in Finnish EFL textbooks for upper secondary schools was similar to that in present-day English from two perspectives: frequency of occurrence and modal meanings. To this end, several corpus-findings representing the modal use in the real world were compared to occurrences within a textbook corpus, which was created by Nordberg himself. For frequency of occurrence, he revealed similarities between the two kinds of language data. For the modal meanings, however, he found that textbooks overemphasized the one-sided meaning for each modal in a way that differed from the present-day English.

In sum, previous studies seem to have been carried out mainly from four perspectives: frequency of occurrence, modal meanings, VPSs, and syntactic surroundings. As a first step for studies focusing on Japanese learner's modal usage, this research will pay special attention to frequency of occurrence and the VPSs where modal verbs occur; that is, the leading research questions are (1) which modal do Japanese learners of English over-/under-use as compared to native English-speakers? and (2) are there any differences in preferences that the modal verbs have for the VPSs between Japanese learners of English and native English-speakers?

Methodology

Corpus under Analysis

The *International Corpus Network of Asian English* (ICNALE), which is one of the largest freely-available corpora of Asian learners' English (Ishikawa, 2013), was adopted for the purpose of this CIA research. This corpus is created with the aim of conducting reliable CIA research; therefore, unsurprisingly, the writing conditions are strictly controlled. Specifically, all the writers wrote "about the same topic within the

same amount of time, and they produced essays of the same length and using the same PC environments and references" (Ishikawa, 2013, p. 94). Hence, results obtained from the corpus could potentially give us a reliable insight into learner language. The ICNALE corpus is composed of four kinds of modules: *Spoken monologue*, *Spoken dialogue*, *Written essays*, and *Edited essays*. Among these, only the *Written essays* module was used for the current study, meaning that the current study will leave analysis regarding the modal usage in spoken contexts open for further research. The *Written essays* module consists of the set of argumentative 200-300 words essays on two types of topic: one is *Is it important for college students to have a part time job?* and the other is *Should smoking be completely banned at all the restaurants in the country?*

In most CIA studies learner language has been compared with native-speaking student language as comparable language data (Granger, 1996; Granger, 2015; Ishikawa, 2013). On the other hand, Granger (1998) and Granger (2015) point out that the language produced by native-speaking experts should also be considered rather than focusing only on native-speaking students since their language may not be what learners should imitate (Aston, 2008). The ICNALE corpus can display three types of native language data: native-speaking students, native-speaking teachers, and native-speaking adults (i.e., native speakers who have no experience teaching English). Among these, it would be reasonable to assume that the language produced by native-speaking teachers corresponds to that of native-speaking experts. Learner's language use is, thus, compared with two types of native language so that we can obtain a clear picture of learner language. The following table illustrates the inside information of the participants in the ICNALE corpus. In it, *JNNS*, *NESS*, and *NEST* refer to Japanese non-native students, native English-speaking students, native English-speaking teachers respectively; I will use these abbreviations to refer to each group of writers henceforth.

	number of writers	*tokens*
JNNS	400	177,253
NESS	100	44,749
NEST	44	19,867

Table 1: Information on the INCALE corpus

We can choose to use either the ICNALE online or the ICNALE download version. The latter version was adopted for this research because it gives access to updated data and allows for analysis of the corpus with other corpus tools that researchers would like to use.

Instrumentation

To reveal and compare frequencies of use of the nine central modals employed by the three groups, *Wordsmith Tools 7.0.* (Smith, 2016) was used; among several features that this program has, the *Concord* feature was used to call up a set of concordances having the modal verbs under analysis, allowing for analysis regarding the VPSs.

Analysis Results

This chapter is devoted to describing how the central nine modals are used by the three groups respectively from the primary perspectives in this research (i.e., frequency of use and VPS).

Frequency analysis

Table 2 summarizes how many times each modal is used by each group. Note that, this research excluded instances comprising the modals that are used grammatically incorrectly from the scope of analysis; therefore, Table 2 shows frequency of grammatically correct use of the modals. The following six cases represent the excluded ones.

(1) Modals are followed by two verbs.
…they would not *be enjoy the meal. (W_JPN_SMK0_337_B1_1)
(2) Modals are followed by inflected forms of verbs.
…many companies and shops can *decreased the payment of employee's wage. (W_JPN_PTJ0_019_B1_2)
(3) Modals are not followed by verbs.
…we should much more *friends in a university. (W_JPN_PTJ0_277_A2_0)
(4) Modals are used in inappropriate verb phrase structures.
Secondly, a part time job *can be improved student's communication skill. (W_JPN_PTJ0_350_A2_0)
(5) Verb tense is incorrect.
If I *did not do part time job, I *could not do these experiences. (W_JPN_PTJ0_062_B1_1)
This sentence is grammatically correct, but guessing from the context, this sentence should be, *If I had not done part time job, I could not have experienced these things*.
(6) The place of *not* is incorrect.
So he and his family will be *not happy. (W_JPN_SMK0_098_A2_0)

In Table 2, the modal verbs have been put in descending order of frequency of use observed in JNNSs' essays. The numbers in parentheses represent the order ranked by frequency of use in each group.

Modal	JNNS	NESS	NEST
can	1755	329 (1)	100 (3)
should	1314	271 (2)	101 (2)
will	719	260 (3)	99 (4)
may	363	67 (5)	55 (5)
must	327	14 (8)	16 (7)
would	154	203 (4)	112 (1)
could	130	38 (6)	23 (6)
might	65	31 (7)	8 (8)
shall	1	2 (9)	1 (9)
TOTAL	4828	1215	515

Table 2: Frequency of use in each group

JNNSs and NESSs share the same tendency for the high frequency modals. Both groups use *can* with the most frequency, followed by *should* and *will*; however, the frequency order of all the other modals except for *shall* does not match each other. For example, NESSs infrequently use *must* as compared to JNNSs while they use *would* more frequently than JNNSs.

Compared to NESTs, gaps in frequency order seem to widely differ as compared to those observed in the comparison with NESSs. NESTs use *would* with the most frequency whereas it is one of the mid-frequency modals in JNNSs' essays. The modal *must* belongs to the low frequency modals in JNNSs' essays whereas it would be categorized into mid-frequency modals in NESSs' essays. Overall, JNSSs often use the modals that NESTs tend to infrequently use in their writing.

To see whether or not these gaps in frequency of use are statistically significant, I performed a log-likelihood test on the data. This testing can better uncover under-/over-used items in experimental corpora relative to reference corpora (Seog & Choi, 2018); this is valid when the sizes of corpora to be analyzed are quite different from each other (Rayson & Garside, 2000; Yae, 2015). The web-based statistical tool which was developed by Paul Rayson was used for a comparison of frequency of use, allowing us to reveal if differences in frequency between two language data are statistically significant with a log-likelihood value. This is freely available at http://ucrel.lancs.ac.uk/llwizard.html.

Modal	Log-Likelihood		Bayes Factor**
can	+	26.25*	13.94
should	+	9.60	-2.71
will	-	23.24*	10.93
may	+	5.96	-6.36
must	+	75.25*	62.94
would	-	231.43*	219.12
could	-	0.61	-11.70
might	-	7.79	-4.52
shall	-	3.04	-9.27
TOTAL	+	0.01	-12.30

*: statistically significant at the level of *p* <.05
**: degrees of evidence against the null hypothesis
Table 3: Result of a log-likelihood test- JNNS vs NESS

Table 3 summarizes the result of a log-likelihood test on a comparison of frequency of modal use between JNNSs and NESSs. Considering Bayes factors, the data identifies that JNNSs overuse *can* and *must* as well as underuse *will* and *would* at a level that is statistically significant.

Modal		Log-Likelihood	Bayes Factor**
can	+	53.23*	41.04
should	+	14.88*	2.69
will	-	3.50	-8.69
may	-	4.03	-8.16
must	+	13.58*	1.39
would	-	184.65*	172.46
could	-	3.66	-8.53
might	-	0.06	-12.13
shall	-	2.03	-10.16
TOTAL	+	1.16	-11.04

*: statistically significant at the level of $p < .05$
**: degrees of evidence against the null hypothesis
Table 4: Result of a log-likelihood test- JNNS vs ENST

Table 4 illustrates the result of a log-likelihood test on the differences in frequency of use between JNNSs and NESTs. According to this table, JNNSs overuse *can*, *should*, *must* as well as underuse *would* relative to NESTs at a level that is statistically significant.

VPS analysis

This paper has already stated that the VPSs to be analyzed will follow the Kennedy's taxonomy as summarized above; however, among the nine VPSs, the central nine modals rarely occur in S6-S9. Specifically, in the whole data there are only two instances occurring in those structures, all of which are produced by ENSSs. Thus, these VPSs are excluded from the scope of analysis because such a small number of instances would not be enough to discuss and generalize results, leading to an extreme interpretation.

Table 5 compares the distribution of the VPSs where each modal occurs in JNNSs' writing with that in NESSs' writing, and results are expressed as a percentage.

	S1		S2		S3		S4		S5	
	JNNS	NESS	JNNS	NESS	JNNS	NESS	JNNS	NESS	JNNS	NESS
can	1.8	2.4	96.4	93.6	1.7	4.0	0.1	0.0	0.1	0.0
could	0.0	2.6	96.2	86.8	2.3	7.9	0.0	0.0	0.8	2.6
may	0.3	0.0	95.0	88.5	3.9	6.5	0.8	3.8	0.0	1.2
might	0.0	1.0	92.3	86.2	6.2	6.9	1.5	0.5	0.0	4.9
will	0.4	0.0	94.0	85.1	4.9	4.5	0.3	1.5	0.1	7.5
would	0.0	0.0	94.2	87.1	3.9	0.0	1.9	0.0	0.0	12.9
shall	0.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0
should	0.0	0.7	62.6	64.2	37.3	33.9	0.0	0.7	0.2	0.4
must	0.0	0.0	92.4	85.7	7.3	14.3	0.0	0.0	0.3	0.0
TOTAL	0.7	1.1	86.3	83.8	12.5	11.9	0.2	1.2	0.1	2.0

Table 5: Distribution of the VPS where modal verbs co-occur- JNNS vs NESS

Overall, both groups use the modals in S2 with the most frequency, followed by S3. What is noteworthy in this data can be that in JNNSs' writing S4 and S5 respectively account for only 0.2% and 0.1% of all the instances, meaning one sixth and one twentieth of the incidences occurring in ENSSs' essays. This implies that JNNSs seem to infrequently use the modal verbs in these structures as compared to NESSs. In addition, the data shows that JNNSs rarely use the modal verbs in S1.

Looking closely at each individual modal, another fact to mention is the difference in the use of the modal *may*, which is the modal occurring in S4 most frequently in NESSs' essays while this case rarely occurs in JNNSs' essays. The differences in the use of *will*, *would*, and *might* between the two groups should be another point to note. In NESSs' essays these modals frequently occur in S5 as compared to the other modals whereas in JNNSs' essays, they seldom occur in the structure, implying that depending on the modals, their preferences for the VPSs may vary.

	S1		S2		S3		S4		S5	
	JNNS	NEST	JNNS	NEST	JNNS	NEST	JNNS	NEST	JNNS	NEST
can	1.8	9.0	96.4	81.0	1.7	10.0	0.1	0.0	0.1	0.0
could	0.0	0.0	96.2	65.2	2.3	34.8	0.0	0.0	0.8	0.0
may	0.3	0.0	95.0	92.9	3.9	6.1	0.8	1.0	0.0	0.0
might	0.0	0.0	92.3	92.0	6.2	1.8	1.5	3.6	0.0	2.7
will	0.4	0.0	94.0	80.0	4.9	12.7	0.4	1.8	0.1	5.5
would	0.0	0.0	94.2	100.0	3.9	0.0	1.9	0.0	0.0	0.0
shall	0.0	0.0	100.0	100.0	0.0	0.0	0.0	0.0	0.0	0.0
should	0.0	1.0	62.6	60.4	37.3	36.6	0.0	2.0	0.2	0.0
must	0.0	0.0	92.4	81.3	7.3	12.5	0.0	6.3	0.3	0.0
TOTAL	0.7	1.9	86.3	81.2	12.5	14.0	0.2	1.7	0.1	1.2

Table 6: Distribution of the VPS where modal verbs co-occur- JNNS vs NEST

According to Table 6, both groups use the modals in S2 with the most frequency, followed by S3. The differences in the use of S4 and S5 between the two groups seem to widely differ as compared to those of the other VPSs. Specifically, NESTs use the modals in S4 8.5 times and in S5 12 times more than JNNSs. These facts that JNNSs infrequently use the modals in S4 and S5 relative to NESTs are identical to the results observed in the comparison between JNNSs and NESSs.

A close look at an individual modal reveals that the modal *could* occurring in S3 accounts for 34.8% of all the instances, meaning the rate is 15 times more than that in JNNSs' writing. Most occurrences of S1 tend towards the use of *can*. Besides, the modal *must* contributes to S4 in NESTs' essays most whereas JNNSs do not produce this modal in the structure. Focusing on the modals co-occurring with S5, we can see that in NESTs' writing the modal occurring in S5 most frequently is *will* whereas this use is rare in JNNSs' essays, again indicating that each modal seems to have different preferences for the VPSs.

Discussion

This CIA research has revealed several points worth discussing. The results did not show any significant difference in overall frequency of use between JNNSs and both of the native speaker groups, which is contrary to what previous studies have shown. Those studies identified the learners' overuse of modal expressions (Aijmer, 2002; McDouall, 2012; Xiao, 2017; Yang, 2018) or underuse of them (Fujimoto, 2019; Seog & Choi, 2018) relative to native speakers. Looking at the modal usage at an individual modal level, however, this current study can highlight some notable differences.

Frequency analysis also uncovered JNNSs' overuse of *can* and *must* as well as underuse of *will* and *would* at a level that is statistically significant relative to both of the native speaker groups. First, it is interesting to note that both of the underused modals are members of so-called epistemic modals. In writing, they have an important role because they are used to not only confer author's judgment or evaluation on truth value to their statements, but also build a relationship with readers (Palmer, 1990; Hu and Li, 2015). Using these expressions appropriately would be, thus, the key factor to make one's writing more readable.

Leech et al. (2009) identified that in written contexts, the modal *would* was used most frequently, followed by *will*. Both modals were the modals that their epistemic meanings (i.e., prediction) were dominant rather than their root meanings (i.e., volition/intention) as exemplified by the Longman Spoken and Written English Corpus (Biber et al., 1999). That is, native English speakers may often use the two modals to convey epistemic modality in writing, leading us to argue that filling in the gaps in the use of the two modals can potentially help bring Japanese learner's modal usage in writing close to the native-like usage; therefore, student's underuse of *will* and *would* is a problem which needs to be addressed.

Two of the modals, *can* and *must*, are overused by JNNSs as compared to both of the native speaker groups. There may be several reasons for this. First, the fact that the modal *can* is overused and produced by learners of English most frequently is identical to what previous studies have identified (Btoosh, 2019; Viana, 2006; Yang, 2018). This might be, therefore, a common trend among learners of English. As is well known, the modal *can* can convey three different meanings: permission, possibility, and ability. These meanings are supposed to be acquired by learners even if they are at an early stage (Coates, 1982; Seog & Choi, 2018). As mentioned earlier, modal verbs are among the most challenging items for learners of English. Hence, it is possible that JNNSs used *can* more frequently than the other modals because it might have been the only modal that they were able to use with confidence.

The learners' overuse of *can* and *must* can be potentially explained by the *One to One Principle* (Anderson, 1984), which is the idea that if learners decide to use one form to express one meaning, they would rely on it and would not use other equivalent expressions. For example, the meaning of *can* regarding possibility could be expressed by other lexical items such as *perhaps* or *probably*. Likewise, equivalent expressions of *must* conveying certainty could be paraphrased using *definitely* or *undoubtedly*. In her study, Aijmer (2002) identified that Swedish learners of English tended to overuse modal verbs to express modality, and she considered that this was partly due to EFL/ESL textbooks overemphasizing modal verbs and

underemphasizing other lexical items to express modality. Considering these findings, one possible interpretation is that JNNSs overuse *must* and *can* because they rely largely on the two modals to express modality that they can convey and ignore other equivalent expressions. At the same time, it should be pointed out that we cannot strongly argue this point yet because the current study focuses on the modal verbs only. Thus, what induces learners' overuse of the two modals should be researched more in the future.

Another aspect of this study which should be pointed out is that the modal *should* was found to be a high frequency modal whereas many previous studies identified that it was a member of the mid-frequency modals in the real world (Kennedy, 2002; Seog & Choi, 2018) It was also revealed that the modal was often used in the passive voice. This fact is not surprising because one of the essay topics in the ICNALE corpus is *Should smoking be completely banned at all the restaurants in the country?* Referring to instances of *should* produced by the three groups, we can easily find that all the groups often borrow the expression from the indication sentence.

Semantically speaking, the modal *should* has the function of giving advice. Writers were required to write essays argumentatively; therefore, we can imagine that they would have used this modal to give advice or made a suggestion so as to persuade readers with their arguments. According to researchers (Biber et al., 1999; Leech, 2004; Leech et al., 2009), depending on genres where the language is produced, its language behavior quite varies, meaning that the frequent use of *should* can be one example of what previous studies have shown.

As for the VPSs, both Japanese students and native English-speakers use the modal verbs in S2 with the most frequency, followed by S3. Focusing on the other VPSs, one thing that is noteworthy is that in JNNSs' essays, there are few occurrences of S1 (0.7%). However, this may not be surprising because such use mainly serves as ellipsis in conversational contexts (Kennedy, 2002), meaning that discussing the reasons for the infrequent use of S1 in writing would not give us a valuable insight. JNNSs rarely use the central nine modals with the progressive aspect (i.e., S4) as well as the perfect aspect (i.e., S5) as compared to the two native English-speaker groups. Although co-occurrences of modal verbs with the progressive aspect are not so frequent (Biber et al., 1999; Kennedy, 2002), this use should not be ignored because this can remove ambiguities in interpretations of modal meanings. According to Palmer (1990), the modal *must* as in "He must come tomorrow" (Palmer, 1990, p. 54) conveys the meaning of obligation as well as expresses speaker's certainty. By changing the sentence into "He must be coming tomorrow" (Palmer, 1990, p. 54), such ambiguity can be removed. Having a knowledge of this usage could be helpful especially for low-level learners because they are more likely to produce semantically ambiguous sentences than high-level learners.

Another fact to note would be that JNNSs infrequently use the modals with the perfect aspect. When referring to past events or actions, modal verbs are usually followed by the perfect aspect. The lack of learners' ability to use the modals with this aspect could mean their failures to express their subjectivity on something that happened in the past. Besides, one of the main roles of the modal verbs co-occurring with this structure is to express past unreal situations. In his article, Bryant (1975) pointed out that this usage was one of the common mistakes among Japanese learners of English,

suggesting that teaching materials or language teachers should carefully teach how to use the modals with the perfect aspect.

Suggestion for further studies

This study left several topics concerning learner's modal usage open. First, this research did not investigate whether or not there were any differences in the use of the modal verbs depending on learner's proficiency levels even though the ICNALE corpus can call up the language produced by a specific level of learners classified by CEFR levels. If future studies could identify typical mistakes which are particular to certain levels of learners, it would lead us to understand learner's developmental sequence.

It can also be worth investigating what induces learner's over-/under-use of certain modals. Aijmer (2002) points out that L2 learner's modal use may be influenced by many factors including learner's cultural value in their L1, textbooks, or teachers. Identifying the most influential factor on learner's modal use can offer a helpful insight into the way to teach them effectively.

Conclusion

The primary questions were (1) which modal do Japanese learners of English over-/under-use as compared to native English-speakers? and (2) are there any differences in preferences that the modal verbs have for the VPSs between Japanese learners of English and native English-speakers? As for the first research question, one of the significant findings was the Japanese students' underuse of the epistemic modals that native speakers would be more likely to use in writing. Given that the language produced by native speakers is the goal that learners should achieve, the underused modals should be more stressed in classrooms. In addition, the results revealed the overuse of *must* and *can* by Japanese learners. Introducing other lexical items having similar functions with the two modals can be helpful so that learners have a wide range of lexical items to reflect their feelings more accurately.

As for the second question, the results indicated that JNNSs tended to use the modal verbs in S2 and S3; they infrequently use the modals in S4 and S5 relative to native English-speakers. This may be one of the problems to combat because a situation is presented from different perspectives and viewpoints, and one verb form cannot describe every situation (Smith, 1983). It may be, however, hard for learners to acquire all the forms that modal verbs can take because they can co-occur with a wide range of VPSs. Fortunately, the results obtained in this study can potentially help propose a solution for this. This research excluded several of the nine VPSs from the scope of analysis because their occurrences were quite rare. Besides, the results showed that depending on the modal verbs, their preferences for the VPSs quite varied. That is, English teachers do not have to teach all the modals in exactly the same way; this information would give English teachers a helpful insight into which modal should be more prioritized in which VPSs.

The current study has shown that CIA research can be valid to obtain a picture of learner's interlanguage for at least the modal usage. Unsurprisingly, the next step would be to investigate how we can fill in gaps between learner language and native

language. One suggestion can lie in the improvement of textbooks because the content and curriculum of English classes in Japan tends to be largely dictated by the contents of the textbooks used (Hino, 1988). Using corpus-based textbooks can expose learners to language patterns occurring in the real world, leading them to acquire the native-like language use.

References

Aijmer, M. (2002). Modality in advanced Swedish learners' written interlanguage. In S. Granger, & J. Hung, & S. Petch-Tyson (Eds.), Computer learner corpora, second language acquisition and foreign language teaching (pp. 55-76). Amsterdam/Philadelphia: John Benjamins.

Anderson, R. W. (1984). The one to one principle of interlanguage construction. Language Learning, 34(4). 77-95.

Aston, G. (2008). It's only human…. In A. Martelli, V. Pulcini (Eds.), Investigating English with Corpora. Studies in Honour of Maria Teresa Prat (pp. 343-354). Monza: Polimetrica International Scientific Publisher.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). Longman grammar of spoken and written English. Harlow: Pearson Education.

Bryant, W. H. (1984). Typical errors in English made by Japanese ESL students. JALT Journal, 6, 1-18.

Btoosh, M. A. (2019). Modals in Arab EFL learners' composition: A corpus-based approach. Linguistics and Literature Studies, 7(3), 100-109.

Celce-Murcia, M., & Larsen-Freeman, D. (1999). The grammar book: An ESL/EFL teacher's course. (2nd ed.). Boston: Heinle & Heinle.

Coates, J. (1983). The semantics of the modal auxiliaries. London: Longman.

Cook, W. A. (1978). Semantic structure of English modals. TESOL Quarterly, 12(1), 5-15.

Decapua, A. (2008). Grammar for teachers: A guide to American English for native and non-native speakers. New Rochelle: Springer.

Depraetere, I., & Reed, S. (2006). Mood and modality in English. In B. Aarts, & A. McMahon (Eds.), The Handbook of English Linguistics (pp. 269-290). Malden: Blackwell Publishing Ltd.

Fujimoto, K. (2019, July). Epistemic modal verbs and adverbs in Japanese university students' academic writing. Paper presented at International Corpus Linguistic Conference (CL2019). Abstract.

Granger, S. (1996). From CA to CIA and back: An integrated contrastive approach to computerized bilingual and learner corpora. In K. Aijmer, B. Altenberg, & M. Johansson (Eds.), Languages in contrast (pp. 37-51). Lund: Lund University Press.

Granger, S. (1998). The computer learner corpus: A versatile new source of data for SLA research. In S. Granger (Eds.), Learner English on computer (pp. 3-18). London & New York: Addison Wesley Longman.

Granger, S. (2015). Contrastive interlanguage analysis: A reappraisal. International Journal of Learner Corpus Research, 1, 7-14.

Hino, N. (1988). Nationalism and English as an international language: The history of English textbooks in Japan. World Englishes, 7, 309-314.

Holmes, J. (1988). Doubt and certainty in ESL textbooks. Applied. Linguistics, 9(1), 21-44.

Hu, C., & Li, X. (2015). Epistemic modality in the argumentative essays of Chinese EFL learners. English Language Teaching, 8(6), 20-31.

Ishikawa, S. (2013). The ICNALE and sophisticated contrastive interlanguage analysis of Asian learners of English. In S. Ishikawa (Eds.), Learner corpus studies in Asia and the world, 1 (pp. 91-118). Kobe: Kobe University.

Kennedy, G. (2002). Variation in the distribution of modal verbs in the British National Corpus. In R. Reppen, S. Fitzmaurica, & D. Biber (Eds.), Using corpora to explore linguistic variation (pp. 73-90). Amsterdam: John Benjamins.

Khojasteh, L., & Kafipour, R. (2012). Have the modal verb phrase structures been well presented in Malaysian English textbooks? English Language and Literature Studies, 2(1), 35-41.

Khojasteh, L., & Reinders, H. (2013). How textbooks (and learners) get it wrong: A corpus study of modal auxiliary verbs. Applied Research on English Language, 2(1), 33-44.

Leech, G. (2004). Meaning and the English verb. (3rd ed.). London: Pearson Education Ltd.

Leech, G., Hundt, M., Hair, C., & Smith, N. (2009). Change in contemporary English. Cambridge: Cambridge University Press.

Mukundan, J., & Khojasteh, L. (2011). Modal auxiliary verbs in prescribed Malaysian English textbooks. English Language Teaching, 4(1), 79-89.

McDouall, A. (2012). A corpus based investigation into the use of English modal auxiliaries by adult Korean L2-learners. KLING, 6, 33-44.

Nordberg, T. (2010). Modality as portrayed in Finnish upper secondary school EFL textbooks: A corpus-based approach. (Unpublished master's thesis). Retrieved from https://helda.helsinki.fi/handle/10138/19357.

Palmer, F. R. (1990). Modality and the English modals. (2nd ed.). New York: Longman.

Römer, U. (2004). A corpus-driven approach to modal auxiliaries and their didactics. In J. Sinclair (Eds), How to use corpora in language teaching (pp. 185-199). Amsterdam: John Benjamins.

Rayson, P., & Garside, R. (2000). Comparing corpora using frequency profiling. In A. Kilgariff & T. B. Sardinha (Eds.), Proceedings of the workshop on comparing corpora- Volume 9 (pp. 1-6). Hong Kong: Association for Computational Linguistics.

Scott, M. (2016) WordSmith Tools version 7 [Computer software]. Stroud: Lexical Analysis Software.

Seog, D. Soon-Young., & Choi, I. (2018). Interlanguage development of young Korean EFL learners' modal usage: A learner corpus study. Linguistic Research, 35, 83-103.

Smith, C.S. (1983). A theory of aspectual choice. The journal of Linguistic Society of America, 59(3), 479-501.

Viana, V. (2006). Modals in Brazilian advanced EFL learners' compositions: A corpus-based investigation. Profile Issues in Teachers' Professional Development, 7, 77-86.

Wilson, A. (2005). Modal verbs in written Indian English: A quantitative and comparative analysis of Kolhapur corpus using correspondence analysis. ICAME Journal, 29, 151-169.

Xiao, Y. (2017). Chinese ELF learners' acquisition of modal verbs: A corpus-based study. International Journal of English Linguistics, 7(6), 164-170.

Yae, S. (2015). Grammaticalization of marginal modal verb ought to: A corpus-based approach. Linguistic Research, 32(3), 773-793.

Yang, X. (2018). A corpus-based study of modal verbs in Chinese learners' academic writing. English Language Teaching, 11(2), 122-130.

Contact email: shusaku-nakayama@outlook.jp