

A Corpus-Based Study of Modal Verbs in English Textbooks for Junior High Schools Used in Japan

TATSUYA KAJIYAMA

Abstract

This paper constructed a textbooks corpus with English textbooks for junior high schools used in Japan (approximately 60,000 words) to see whether or not the frequencies of modal auxiliary verbs in each grade and in each modal auxiliary verb were statistically different. The Friedman test was applied to answer these questions. As a result, there were statistically significant differences in general among the grades and among modal auxiliary verbs while the post hoc comparisons did not show significance. This could be because of Bonferroni's correction. Even though the result of the comparison was not like our expectation, it could be inferred that there were differences between Grade 2 and Grade 1 and 3 as well as there were differences between “can”, “will”, and the others.

Keywords: English learning, modal auxiliary verbs, pedagogical corpus

1. Introduction

In recent years, corpus-based textbook studies have been carried out because the textbook is an essential material for language learners (e.g., Bouhlal et al., 2018; Khojasteh & Reinders, 2013; McEnery & Kifle, 2002; Mukundan & Khojasteh, 2011; Römer, 2004). Some studies examined modal auxiliary verbs because they are one of the difficult grammatical

aspects to learn owing to their forms and meanings. As Celce-Murcia and Larsen-Freeman (1999) pointed out, unlike verbs, auxiliary verbs do not take a third person singular “-s”, and their present-past forms do not indicate present-past tense. Moreover, they are not followed by infinitives like “can go”. Furthermore, they generally have two meanings. Coates (1983) categorized them into root modality and epistemic modality. Root modality could be defined as modality representing “a logic of obligation and permission” (Coates, 1983, p. 21) On the other hand, epistemic modality was “concerned with the speaker’s assumption or assessment of possibilities” (Coates, 1983, p. 18). Palmer (2001) categorized the two meanings into propositional modality and event modality. The former meant “the speaker’s attitude to the truth-value or factual status of the proposition” (Palmer, 2001, p. 24) while the latter referred “to events that are not actualized” (Palmer, 2001, p. 70). Root modality corresponded to event modality and epistemic modality did to propositional modality.¹ Because of these complex features, English modal auxiliary verbs are difficult to learn for English learners.

We turn now to the issue of Japanese English learners. A few studies have been conducted about the frequencies and usages of modal auxiliary verbs in Japanese learners. Ishida (2011) surveyed English essays written by Japanese university learners. The author ascertained the discrepancies between the Japanese learners and the English native speakers about the frequencies and usages of modal auxiliary verbs. Nakayama (2020), furthermore, explored the misuse of modal auxiliary verbs by Japanese learners and suggested the necessity of future investigations of the reasons for the misuse.

While there were a few studies on how accurately Japanese English learners could use modal auxiliary verbs with corpus tools, none of the studies have been done to explore modal auxiliary verbs used in English

textbooks for junior high schools. The purpose of this paper is to examine the frequencies of modal auxiliary verbs in English textbooks for junior high schools used in Japan because the amount of input would influence on proficiencies of Japanese learners and the textbooks would be central input for them to learn English.

This paper consists of five sections. In section 2, we will introduce previous corpus studies as well as studies for Japanese English learning. In section 3, we will explain the way of corpus research in this paper. The research results will be demonstrated in the section 4. We will discuss them in the following section, and finally comment limitations and conclusion to this study.

2. Previous Corpus-Based Studies

2.1 Modals in ESL/EFL Textbooks

Several studies investigated how textbooks dealt with modal auxiliary verbs. Hyland (1994) established a corpus of 22 textbooks for English as Academic Purpose (EAP) and English as Specific Purpose (ESP) to see how these textbooks dealt with hedging devices. In this study, modal verbs, lexical verbs, modal adverbs, modal adjectives, and modal nouns were the research targets. According to Hyland (1994), “the overall picture indicates a need for greater and more systematic attention to be given this important interpersonal strategy” (p. 246). Furthermore, even “the fact that modals are the most identified and widely used means of hedging in academic writing” (Hyland, 1994, p. 247), most textbooks took a limited page to explain the modal auxiliary verbs. This was probably problematic because “anything not included can be safely omitted from a course” (Hyland, 1994, p. 252). It can be said that this study not only showed the issues concerning the 22 ESP and EAP textbooks but also implied the importance of textbooks research from the perspective of frequency.

Römer (2004) compared modal auxiliary verbs in German-English textbooks and the British National Corpus (BNC) and found that German textbooks introduced “can” at first, followed by “must” and “may”, while the most frequent word in the BNC was “will” followed by “would”. Furthermore, the author indicated that the German textbooks had to use more passive and interrogative construction for “should”, “must”, “may”, and “might”. Using these findings, Römer suggested the necessity that the German textbooks should follow the information from the native speaker corpus.

Khojasteh and Reinder (2013) also investigated how the Malaysian textbooks were different from the BNC in terms of the frequencies of modal auxiliary verbs. This study shed light on the relationship between Malaysian textbooks and Malaysian learners. They studied how much the learners learn from the textbooks. Considering the frequencies of modal auxiliary verbs, “the order of frequency of most modals in the Malaysian textbooks does not match that found in native speaker corpora” (Khojasteh & Reinder, 2013, p. 40). The authors also found that the learners’ production of modal auxiliary verbs seemed not to be coincident with the presentation in the textbooks. In this way, Khojasteh and Reinder investigated the frequencies of modal auxiliary verbs among the Malaysian textbooks, the BNC, and the Malaysian learners.

Bouhlal, Horst, and Martini (2018) compared the frequencies and the meanings of modal auxiliary verbs in English as Second Language (ESL) textbooks to those of native speakers of English. The results showed that ESL textbooks used modal auxiliary verbs more often than the native speakers of English, except “could”, “might”, and “shall”. Moreover, the study conducted semantic analyses on each modal auxiliary verb. It found, for example, that ESL textbooks included “should” whose meaning was “suggestion” more often than the native speakers’ corpus owing to the data including “instructions” sections.

Up to this point, we have overviewed the literature which surveyed the English textbooks published in other countries. The main point of the literature was that the frequencies in the textbooks were dissimilar to those in the English native speaker corpus. Moreover, the literature showed that the corpus-based study could be available to analyze the textbooks. Let us now turn to the studies of Japanese English learning.

2.2 Previous studies of Japanese English learning

Ishida (2011) investigated to see how university students in Japan used the English modal auxiliary verbs. The author collected English essays from the 140 Japanese students (27,855 words in total) and compared them with Louvain Corpus of Native English Essays (LOCNESS). This study found that Japanese learners tended to overuse “can”, “may”, and “must” and underuse “could” and “would”. Furthermore, the upper-level students in this study significantly underused “will” and “would” and overused “must”. On the other hand, the lower-level students significantly underused “would”, “should”, and “could” and overused “might” and “must”. In this fashion, Ishida discovered that the Japanese learners could not use modal auxiliary verbs in the same way as the English native speakers.

Nakayama (2020) also argued that the Japanese learners used English modal auxiliary verbs differently than native speakers of English. He noted that the Japanese underused “will” and “would” and overused “can” and “must”. Furthermore, he focused on the aspects that the Japanese tended to use the progressive and perfective aspects less often than the English native speakers. This study also showed the fact that the Japanese learners could not use modal auxiliary verbs properly. Nakayama emphasized the importance of future studies to determine the factors influencing Japanese learners.

Muraoka (2010) investigated the type frequencies of each

part-of-speech in Japanese English textbooks published in 2005. However, the token frequencies should have been counted if the research goal was to investigate the frequencies of modal auxiliary verbs in each textbook. Furthermore, Muraoka did not examine the token frequencies of each modal auxiliary verb in each textbook. Because of these reasons, differences between each textbook regarding the token frequencies need to be examined.

Aizawa and Harada (2015) investigated how English grammar books and workbooks dealt with three types of modal auxiliary verbs (“will”, “can”, and “must”) and their correspondent quasi-modal auxiliary verbs (“be going to”, “be able to”, “have to”). Moreover, the authors conducted research on the degree of comprehension of Japanese university students about these six types of modals. The research reported that the degree of comprehension was quite low and implied the possibility that the very simplified explanation about the modals in the grammar books and workbooks might be a factor for that comprehension.

As above, a few studies have been conducted on Japanese English learners and grammar books. However, it remains to be tested to see how English textbooks for junior high schools used in Japan dealt with modal auxiliary verbs.

2.3 Rationale of the Present Study

Many researchers applied a corpus-based analysis to investigate English textbooks. The foci of such research were to determine the discrepancies between materials and actual language use. Considering English learning in Japan, some researchers examined the comprehension of Japanese learners. Furthermore, a few analyses were carried out to see how Japanese English textbooks and grammar books dealt with modal auxiliary verbs. However, there are still few, if any, studies of English

textbooks in regard to modal auxiliary verbs. In specific, little is known about the frequencies of modal auxiliary verbs in English textbooks used in Japan. In addition, as Aizawa and Harada (2015) indicated the possibility that the Japanese learners have not developed since secondary school, it would be necessary to investigate the English textbooks for junior high schools used in Japan. These facts motivated the author to investigate the junior high school English textbooks in Japan and to find an aspect of learning English modal auxiliary verbs of the Japanese learners.

3. Method

3.1 Research Questions

Based on the previous research findings, this study aims to answer the two research questions:

- (1) Are there any differences in the frequencies of modal auxiliary verbs between grades?
- (2) Are there any differences in the frequencies of each modal auxiliary verb?

3.2 Corpus

A total of 18 junior high school English textbooks published in 2016² (one textbook for each grade from each of six publishers) in Japan was chosen for the target materials. Table 1 shows a list of them. These textbooks were approved by the Ministry of Education, Culture, Sports, Science and Technology (MEXT). They followed a communicative approach and typically included passages for readings, practices, grammar explanations, projects, and tasks. The passages in each lesson and supplementary reading materials were extracted to build a corpus. The basic data of textbooks are in Table 2 and Table 3.

Table 1*List of Junior High School English Textbooks*

Textbook Name	Publisher	Year
Columbus 21 English Course 1–3	Mitsumura Toshō	2016
New Crown English Series New Edition1–3	Sanseido	2016
New Horizon English Course 1–3	Tokyo Shoseki	2016
One World English Course 1–3	Kyoiku Shuppan	2016
Sunshine English Course1–3	Kairyudo	2016
Total English 1–3	Gakko Toshō	2016

Table 2*Details of the Japanese English Textbooks for Junior High Schools Corpus*

Number of textbooks	18
Number of words	56,579
Average Types	806
Average Tokens	3,143
Average STTR	38.69

Table 3*Statistics of Vocabulary in English Textbooks for Junior High Schools*

Publishers	Types	Tokens	STTR	Publishers	Types	Tokens	STTR
A–1	493	1895	32.79	D–1	485	1764	34.51
A–2	796	3592	36.41	D–2	935	3599	40.12
A–3	1200	5689	38.28	D–3	1207	4489	42.98
B–1	525	1820	35.54	E–1	442	1487	37.87
B–2	879	3408	41.44	E–2	816	3243	40.82
B–3	1179	4827	40.82	E–3	993	4133	42.89
C–1	480	1747	35.18	F–1	525	1777	36.79
C–2	796	2943	37.86	F–2	815	2992	38.31
C–3	1051	3996	39.99	F–3	891	3178	43.76

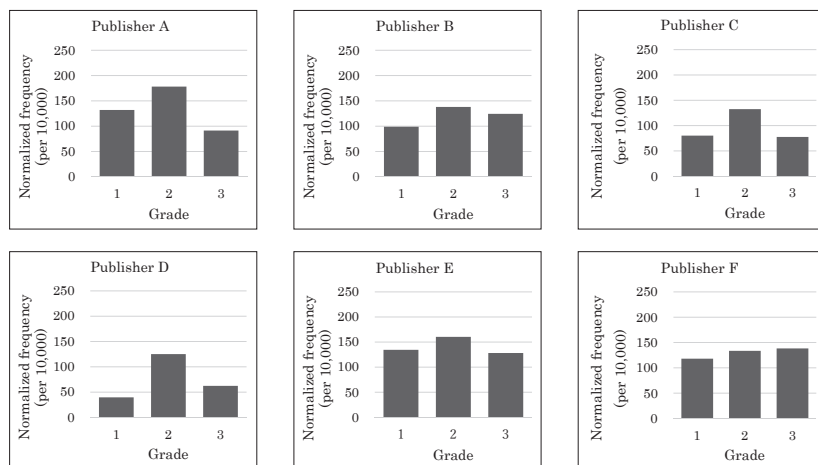
Note. The publishers' names were anonymized. The File Information function in the CasualConc software was used to calculate the lexical frequencies. STTR stands for standardized type/token ratio; M and SD stand for mean and standard deviation respectively. A-1 means, for example, Grade 1 textbook of Publisher A.

3.3 Procedure and Data Analysis

The textbooks were scanned and processed with Optical Character Recognition (OCR) software. Only the passages in each lesson and supplementary reading materials were extracted and encoded with the UTF-8. The data of the readings were converted into text files (.txt). The author examined all of the data to modify typographical errors due to the incompleteness of the OCR technology. At the final step, the data were imported into the CasualConc software (Version 2.0.7) (Imao, 2017) to build a corpus.

Figure 1

Bar Chart of the Normalized Frequencies of Nine Modal Auxiliary Verbs



The main nine modal auxiliary verbs were chosen from the corpus: “can”, “could”, “may”, “might”, “shall”, “should”, “will”, “would”, and “must” (Quirk et al., 1985), and the absolute numbers were counted. They were normalized by 10,000 because the maximum number of tokens was 5,689 of Grade 3 of Publisher A. The total normalized frequencies of the modal auxiliary verbs are described in the Bar Chart in Figure 1.

4. Results

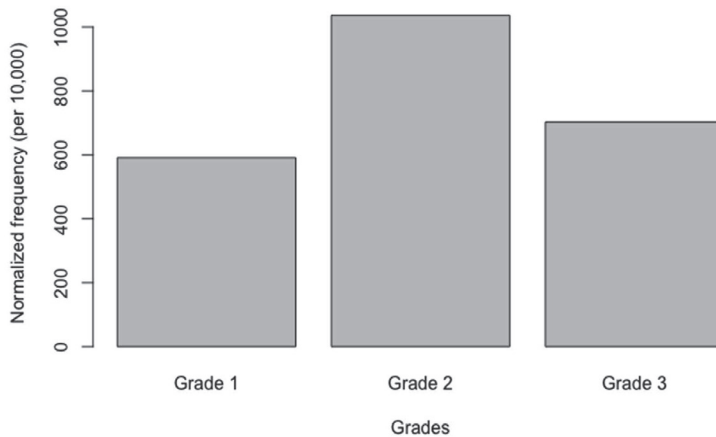
Table 2 and 3 show the detailed information about the textbook corpus. The average types, tokens, and STTR were 806, 3,143, and 38.69 respectively. Publishers A and B had the top two tokens. These were followed by Publisher D whereas Publishers C and E had nearly equal tokens. Publisher F had the least.

Figure 1 shows that the total frequencies of Publisher B for Grades 2 and 3 are almost the same. This is because the textbook of Publisher B for Grade 3 includes the examples of “could” more frequently than any other Publishers for Grade 3. The total number of textbooks of Publisher F for Grade 2 is less frequent than that for Grade 3. This would be attributed to the fact that it includes relatively small input about the examples of “can”. The other four publishers’ textbooks for Grade 2 have larger numbers than Grade 1 and Grade 3. This probably indicates that Japanese learners would gain more input about modal auxiliary verbs in Grade 2 than in Grades 1 and 3.

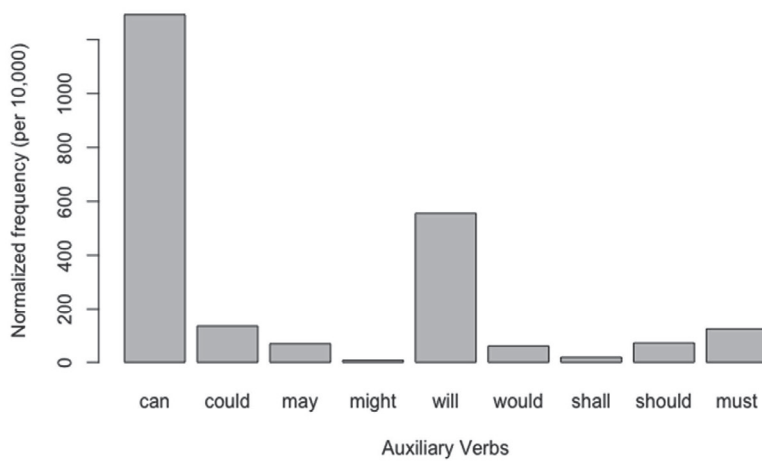
Now, we turn to statistical analysis to answer the RQs. Figure 2 shows the total frequencies of modal auxiliary verbs in each Grade.

Figure 2

Bar Chart of Total Frequencies of Modal Auxiliary Verbs in each Grade

**Figure 3**

Bar Chart of Total Frequencies of Nine Modal Auxiliary Verbs



This implies there would be differences between Grade 1 and 2 and Grade 2 and 3. Figure 3 shows the total frequencies of nine modal auxiliary verbs. This indicates that the frequency of “can” is the highest followed by “will”. The frequencies of “could”, “may”, “would”, “should”, and “must” could be almost the same. The frequencies of “might” and “shall” are close to 0. From Figure 3, there may be statistically significant differences among “can”, “will”, and the others.

To analyze differences in Grades and modal auxiliary verbs, the Friedman test which was a nonparametric test used for a one-way factor was applied. The results are shown in Table 4. There were statistically significant differences in the frequencies of Grades and modal auxiliary verbs. Furthermore, their effect sizes were quite large.

For post hoc comparison, the Wilcoxon signed-rank test was performed for Grades and modal auxiliary verbs respectively. The p values were adjusted by Bonferroni’s correction. As a result, unexpectedly, there was no significantly different pair among Grades and among modal auxiliary verbs. This result will be discussed in the next section.

Table 4

The Results of Friedman Test for Grades and Modal Auxiliary Verbs³

Factors	df	stat.	p	Kendall’s W
Grade	2	10.333	.006	.861
Aux	8	41.799	<.001	.871

Note. For two-factor nonparametric measures, the Scheirer-Ray-Hare test might be adequate. The test could also calculate interaction between factors. However, the test required more than five observations in each cell, but the data in the study did not satisfy the condition. In addition, interaction was not a central interest here. Because of these reasons, this study applied the Friedman test. $W >.50$ means large effect size. (Cohen, 1988).

5. Discussion

In this section, we will discuss the statistical results above. First, it was confirmed that there were statistically significant differences among the three Grades, and its effect size was large. Additionally, Figure 2 indicated that Grade 2 included the highest frequency in the Grade. The Friedman test would express this fact. However, unexpectedly, there was no pair showing the statistically significant difference: $p = .094$ (Grade 1 and Grade 2), $p = .468$ (Grade 1 and Grade 3), and $p = .094$ (Grade 2 and Grade 3), while the effect size was large. This could be because Bonferroni's correction was applied for the post hoc comparison. The correction modified a significant level based on the number of pairs. According to Brezina (2018), "Bonferroni correction is fairly strict and hence the test is fairly conservative" (p. 194). Due to this, we could not statistically find the differences among Grades. In brief, it seems reasonable to conclude that there could possibly be differences between Grade 2 and Grade 1 and 3 from Figure 2 and the Friedman test even though the post hoc comparison did not show the result as expected.

Second, it was verified from Table 4 that overall, there were statistically significant differences among modal auxiliary verbs, and its effect size was large. From Figure 3, this result would be based on the frequencies of "can" and "will". However, the question which has to be discussed was the result of the post hoc comparison. There was no significantly different pair among modal auxiliary verbs. This was also due to Bonferroni's correction. In the case of modal auxiliary verbs, probabilities were multiplied by one-thirty-sixth. This correction might probably lead to the complicated result that no pair was significantly different although the overall effect size was large. To sum up, it can be inferred that Figure 3 and Table 4 would show significant differences about the frequencies among "can", "will", and the other modal auxiliary verbs in spite of the post hoc

comparison.

As above, this study has analyzed the frequencies in each Grade and in each modal auxiliary verb. Even though the post hoc comparisons did not show a significant effect despite our expectation, we could probably achieve some positive results from Figure 3 and 4 and the Friedman test. Next, let us now turn to the limitations of this study.

6. Limitations

Firstly, this study did not closely investigate the meanings of the modal auxiliary verbs. This may be highly problematic because, as Celce-Murcia and Larsen-Freeman (1999) claim, they have at least two meanings, and only frequency could not explain the whole figures. For example, “will”, which had the second-highest frequency, not only has the meaning of “volitive” but also shows a future tense marker. Likewise, the other modal auxiliary verbs have more than two meanings. It is impossible to understand entire figures unless the meanings are taken into consideration. Future research has to address this issue.

Second, the newest textbooks have to be analyzed in the same way. As mentioned in the note, this study analyzed the textbooks published in 2016 although the newest junior high school textbooks were published in 2021. Especially, the new version of textbooks will include examples of the subjunctive mood. (Ministry of Education, Culture, Sports, Science and Technology, 2017, p. 51). From this revision, the examples of the hypothetical meaning of “would”, “could”, and “might” may increase. The future study should focus on the newest ones and compare them with the results of this study.

Lastly, this study examined only the junior high school textbooks, but the high school textbooks should be also analyzed. The high school textbooks should have more frequent input than the junior high school

textbooks. This probably means that the Japanese learners are influenced more by them. Future studies must address this issue.

7. Conclusion

This corpus-based study addressed the two issues related to the frequencies of modal auxiliary verbs in the junior high school English textbooks in Japan. It showed that the corpus-based textbook analysis was valid to obtain and interpret the data from the textbooks. The findings would be helpful for designers of textbooks and language teachers. At the next step, the newest textbooks and the high school English textbooks should be investigated. These would be a rich resource for people involving English learning in Japan.

Notes

1. Although there are differences between the definitions, it is not our purpose to speculate them.
2. Although new version of the junior high school textbooks has been used since 2021, the findings of this study are based on the textbooks published in 2016.
3. The statistics was calculated with R (Version 4.1.2) and R Studio (Version 1.3.1093).

References

- Aizawa, K., & Harada, Y. (2015). Bunpokyouzai ni okeru houjyodoushi no teijihou to gakushusha no rikai [Presentation of modals in grammar-based reference books and learners' comprehension]. *Kyozaigaku Kenkyu*, 26, 33–40. https://doi.org/10.18972/kyozai.26.0_33
- Bouhlal, F., Horst, M., & Martini, J. (2018). Modality in ESL textbooks; Insights from a contrastive corpus-based analysis. *The Canadian Language Review*, 74(2), 227–252. <https://doi.org/10.3138/cmlr.3075>

- Brezina, V. (2018). *Statistics in corpus linguistics*. Cambridge University Press. <https://doi.org/10.1017/9781316410899>
- Celce-Murcia, M., & Larsen-Freeman, D. (1999). *The grammar book: An ESL/EFL teacher's course* (2nd ed.). Heinle & Heinle Publishers.
- Coates, J. (1983). *The semantics of the modal auxiliary verbs*. Croom Helm.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum. <https://doi.org/10.4324/9780203771587>
- Hylland, K. (1994). Hedging in academic writing and EAF textbooks. *English for Specific Purposes*, 13(3), 239–256.
- Imao, Y. (2017). *CasualConc* (Version 2.0.7). <https://sites.google.com/site/casualconcj/>
- Ishida, T. (2011). Nihonjin eigo gakushusha corpus wo motiita houjyoudoushi no shiyou ni kansuru kenkyu [The Use of Modal Auxiliaries in Japanese EFL Learners' Corpus]. *The JACET Chubu Journal*, 9, 57–73.
- 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. <http://dx.doi.org/10.22108/are.2013.15462>
- McEnery, T., & Kifle, N. A. (2002). Epistemic modality in argumentative essays of second-language writers. In J. Flowerdew (Eds.), *Academic Discourse* (pp. 182–195). Longman.
- Ministry of Education, Culture, Sports, Science and Technology. (2017). Chūgakko gakushū shidou youryou [The course of study for secondary school]. https://www.mext.go.jp/content/20210531-mxt_kyoiku01-100002608_010.pdf
- Mukundan, J., & Khojasteh, L. (2011). Modal auxiliary verbs in prescribed Malaysian English textbooks. *English Language Teaching*, 4(1), 79–89. 10.5539/elt.v4n1p79

- Muraoka, R. (2010). Chūgakkō kenteikyōkasho de gakushūsareru goi, gakushūsharenai goi [Learned and non-learned vocabulary in the secondary school textbooks]. *Eiken Bulletin*, 22, 182 – 203.
- Nakayama, S. (2020). Contrastive interlanguage analysis of modal auxiliary verbs usage by Japanese learners of English in argumentative essays. In The International Academic Forum (Eds.), *The IAFOR international conference on education – Hawaii 2020 official conference proceedings* (pp. 507–522). IAFOR. <https://papers.iafor.org/proceedings/conference-proceedings-iicehawaii2020/>
- Palmer, F. R. (2001). *Mood and modality*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139167178>
- Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of the English language*. Longman.
- Römer, U. (2004). A corpus-driven approach to modal auxiliaries and their didactics. In J. M. Sinclair (Eds.), *How to use corpora in language teaching* (pp. 185–199). John Benjamins. <https://doi.org/10.1075/scl.12>