LingPoet: Journal of Linguistics and Literary Research

Modal Verbs in Spoken and Written Text of BNC Sampler: Corpus-Based Study

Winda Ayuanda[1], Dindasari Sidabalok[2], Luthvia Chairunnisa[3] and T. Silvana Sinar[4]

[1,2,3,4]*Universitas Sumatera Utara*

Abstract. The purpose of this study is to find out which type of text between written and spoken text that has the highest frequency of modal usage. It analyzes the personal pronouns which be followed by modal verbs. This study employed a corpus linguistic method, relying on quantitative and qualitative techniques. In quantitative design, the researchers examined the frequencies of modal verbs used in spoken and written text. In the qualitative design, the researchers analyzed the functions of modal verbs in representing the expression shown from the text. The data source of this study is a corpus named BNC (British National Corpus) Sampler. The corpus has 184 texts with total amount of word tokens in all corpus texts is about two million. The result show the tendencies of spoken text are mostly to express permission, possibility, and ability. Meanwhile the written text mostly tends to express obligation, and necessity. Also there are similarities and differences between spoken text and written text. The similar finding is the personal pronoun that appears the most and the least are *I* and *she*. The different finding is the second most frequent pronoun which appears in spoken text is *you*, meanwhile in written text is *it*.

1   Introduction

In understanding a language, it cannot be separated from the elements that construct the language. One of the important parts to be discussed is grammar. In English, grammatical components are not only subject, predicate (verb) and compliments. Besides that, an understanding of modal auxiliary verbs is also an important focus to be studied, not only in spoken text, but also in written text. Winiharti, 2012 refers to several expert views regarding to modal verbs, Saeed, 2003 argues that modality is a cover term for devices which allow speakers to express varying degrees of commitment to, or belief in, a proposition. He also claims that modality is one important semantic category which operates at the sentence level. In other words, modality is a speaker's attitude toward the content of the speak event or utterance. It is

---

*Corresponding author at: Universitas Sumatera Utara

E-mail address: winda.ayuanda@yahoo.com

usually expressed in modal verbs such as must, have to, should, can, may and will. Correspondingly, Kreidler (1998) notes that modality is the expression of necessity, possibility and probability, and is frequently expressed through modal verbs. According to him, the modal verbs include: can, could, may, might, will, would, must, should, ought, need, have to, have got to.

The purpose of this study is to find out which type of text between written and spoken text that has the highest frequency of modal usage. It also analyzes the personal pronouns which be followed by modal verbs. The limit of this study is analyzing the frequency of nine modal verbs (*shall, can, may, must, could, will, might, should,* and *would*) which will show the different characteristic of spoken and written text, whether the text tends to show ability, permission, volition, prediction, obligation or necessity. The analysis of personal pronoun which be followed by modal verbs also be limited to first person subjective personal pronoun (*I, we*), second person personal pronoun (*you*), third person subjective personal pronoun (*he, she, they*), third person singular neuter personal pronoun (*it*), and indefinite pronoun (e.g. *anyone, everything, nobody, one, none*). The purpose of this analysis is indicating to what pronoun is the aim of text most directed to.

The BNC (British National Corpus) Sampler consists of two collections of written and spoken material of about one million words each, originally compiled to mirror the composition of the full BNC as far as possible. The information of BNC Sampler from the CQPWeb – a tool the researchers use for this analysis – proclaims that this corpus is likely to be of interest to: 1) those who wish to use a smaller corpus with equal amounts of written and spoken material; 2) those who want a resource with highly detailed and accurate part-of-speech annotation. Since the BNC Sampler has equal amounts of written and spoken material, this corpus will be compatible to discuss the frequency of modal auxiliary verbs occur in spoken and written text.

2 Literature Review

Previous research on these modal verbs has been carried out by several researchers, such as, Samodra, & Barly (2022), they collected data from the Corpus of Contemporary American English (COCA) and analysed using Hanks' (2004) Corpus Pattern Analysis technique. From this research, the results showed that *shall* was rare in COCA's academic and fiction registers as the overall frequencies were 59.77 and 68.34 words per million, respectively. From all the 400 tokens being analysed, the researchers found that *shall* in the observed data could be classified as rules and regulations, direction, prediction, volition, and etc. The uses of shall in both domains in COCA varied syntactically, semantically, and stylistically. Oktavianti (2019), the analysis shows that core modal verbs expressing volition and prediction are most frequently found in English, with the modal verb will being the most frequently used modal verb and its preterite form occupies the third position. As with English, Indonesian also considers the

expression of volition and prediction salient because it most frequently uses the modality expression akan. The modal verbs can and could are also used intensively by English speakers, showing that speakers consider the concepts of possibility and permission prominent, as well as prediction and volition. The contrastive analysis shows that, among English and Indonesian speakers, prediction and volition are important notional states in terms of modality expression. Sujatna, Eva and Kasno (2019) from the 6,008,840 data, it is found the verb *would* (1,301,269 data) is the most frequently used, it is followed by the verb *can* (1,228,608 data), the verb *will* (1,118,515 data), the verb *could* (885,835 data), the verb *may* (494,144 data), the verb *should* (435,903 data), the verb *might* (300,258 data), the verb *must* (223,936 data), and the verb *shall* (20,372 data). This study has also revealed the number of elements following each of modal auxiliary verbs found in the COCA. From the data analyzed, it is also reported that the modal auxiliary verb *would* is not only has the highest number in data (as the highest frequency) but also has various number of elements follow the verb. Rizvic-Eminovic & Šukalić (2019), the study focuses on the nine English central modal verbs (can, could, will, would, may, might, shall, should, and must) across the two chosen genres of the COCA corpus – Spoken and Academic genres, which show the greatest number of differences in terms of genre characteristics, such as formality or intended audience. The results suggest that genre can indeed be indicative of the frequency and use of modal verbs, which may be ascribed to two language processes colloquialization and democratization

3   Method

This study focused on the use of modal auxiliary verbs, their function, occurrence, and frequency used in spoken and written text. This study employed a corpus linguistic method, relying on quantitative and qualitative techniques. In quantitative design, the researchers examined the frequencies of modal verbs used in spoken and written text. In the qualitative design, the researchers analyzed the functions of modal verbs in representing the expression shown from the text. The corpus-based approach is used by a tool called CQPWeb.

The data source of this study is a corpus named BNC (British National Corpus) Sampler. The corpus has 184 texts with total amount of word tokens in all corpus texts is about two milion. The corpus was compiled by Oxford University Computing Services. The data obtained in this study are limited on the clauses containing modal auxiliary verbs.

The modal auxiliary verbs to be analized are *shall, can, may, must, could, will, might, should,* and *would*. The personal pronouns which occur before modal verbs to be analyzed are first person subjective personal pronoun (*I, we*), second person personal pronoun (*you*), third person subjective personal pronoun (*he, she, they*), and third person singular neuter personal pronoun (*it*).

After collecting the data from BNC Sampler, the researchers identifed and analized the data. Then, the frequency of each modal verbs occur in both spoken and written text is counted and served in chart to show their percentages. The structure of pronoun + modal verb is also analized, with the findings what pronoun – which occurs before modal verb – is most frequently used in both spoken and written text.

4 Result and Discussion

4.1 Frequency of modal verbs

This section discusses the occurrences of modal verbs in spoken and written text. As presented in Figure 1, suggests that the spoken text is predominantly using modal verbs with percentage 59%, meanwhile the written text only own 41%. For further details of frequency data can be seen from Table 1.
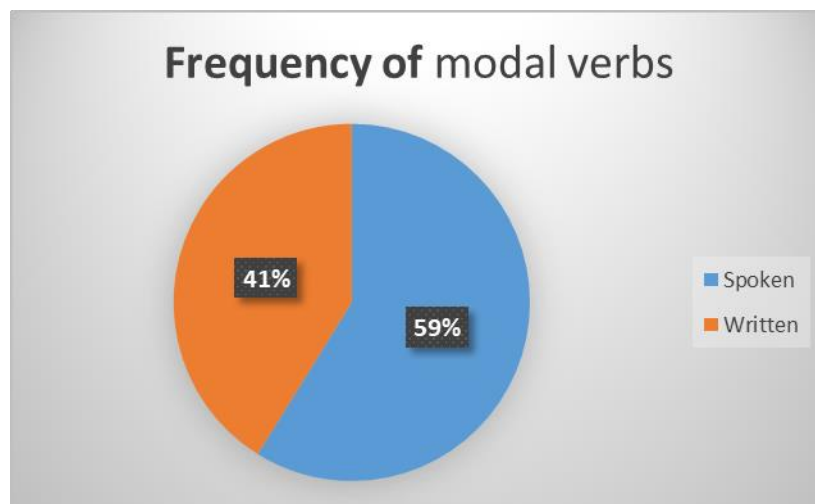

Frequency of modal verbs

41%
59%

■ Spoken
■ Written

Figure 1

Table 1

Kind of Text	Frequency of the usage of modal verbs
Spoken	17,209
Written	12,035

Modal verbs are classified based on their meanings. The researchers have pulled the data of each modals from BNC Sampler corpus using CQPWeb. There are several modal verbs are most frequent in spoken texts, they are *shall, can, could, might,* and *would*. There are also several modal verbs are most frequent in written text, they are *may, must, will,* and *should*. It can be seen from Figure 2. For further detail of total amount of each modal frequency, shown in table 2.
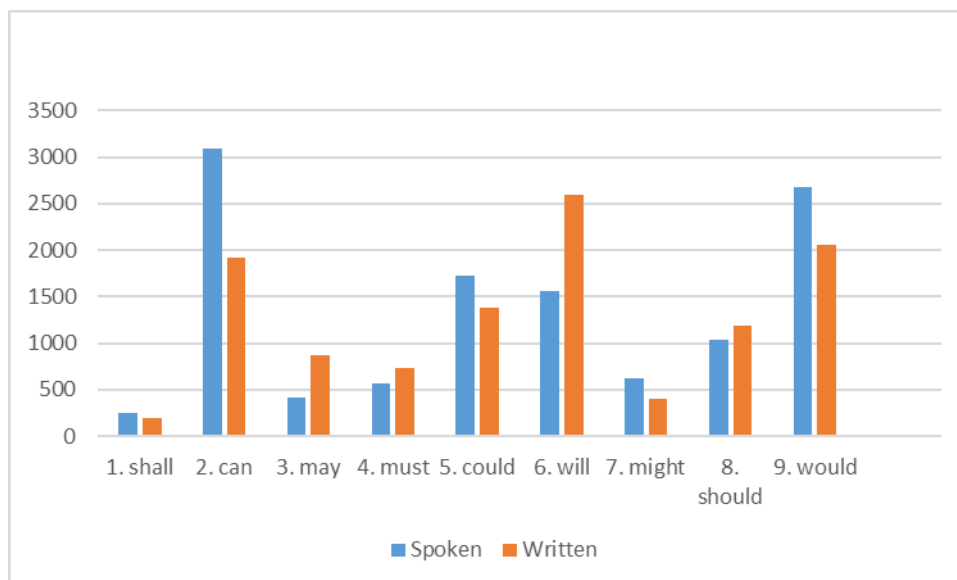
Figure 2

Table 2

freq	shall	can	may	Must	could	will	Might	should	would
Spoken	255	3,089	412	570	1,721	1,557	618	1,037	2,674
Written	191	1,917	875	729	1,376	2,587	400	1,187	2,061

Here is the meaning of each modal qouted from Oxford Advanced Learner's Dictionary 7ᵗʰ edition (Wehmeier & Hornby, 2005), along with its explanation of the domination whether it used most in spoken text or written text.

Shall

Used with *I* and *we* for talking about or predicting the future; used in questions with *I* and *we* for making offers or suggestions, or asking advice; used to show that you are determined, or to give an order or instruction.

The modal verb *shall* is the least used in both spoken and written text. That is because *shall* is old-fashioned word which is passing out of current use. *Shall* is not used very much nowadays. Seeing the few functions of *shall* above, it is relevant that this modal verb is least used. Still, from BNC Sampler corpus, *shall* is used more dominantly in spoken text than written text.

Can

Used to say that it is possible for somebody/something to do something, or for something to happen; used to say that somebody knows how to do something; used with the verbs 'feel', 'hear', 'see', 'smell', 'taste'; used to show that somebody is allowed to do something; used to ask permission to do something; used to ask somebody to help you; used in the negative for saying that you are sure something is not true; used to express doubt or surprise; used to say what somebody/something is often like; used to make suggestions; used to say that somebody must do something, usually when you are angry.

The modal verb *can* is the mostly used in spoken text of all over the other modal verbs. Seeing the many functions of *can* above, it is relevant that this modal verb is mostly used. The written text is also used this modal verb a lot, but the dominant text uses it is spoken text.

May

Used to say that something is possible; used to ask for or give permission; used when admitting that something is true before introducing another point, argument, etc.; used as a polite way of making a comment, asking a question, etc.; used to express wishes and hopes; used to say what the purpose of something is.

The modal verb *may* is not used very much in spoken text. As seen in Figure 2, the comparison of the usage of this verb between spoken and written text is very visible. The verb *may* appears frequently in written text with the total amount of frequency 875, whereas the frequency of its occurance in spoken text is 412. It is the most contrasting comparison among the other modal verbs.

Must

Used to say that something is necessary or very important (sometimes involving a rule or a law); used to say that something is likely or logical; used to recommend that somebody does something because you think it is a good idea.

The modal verb *must* is also mostly used in written text. It is relevant with its functions which is to convey that something is necessary or very important, such as a rule or a law, as we know they are more appropriate delivered by written text.

Could

Used as the past tense of 'can'; used to ask if you can do something; used to politely ask somebody to do something for you; used to show that something is or might be possible; used to suggest something; used to show that you are annoyed that somebody did not do something; used to emphasize how strongly you want to express your feelings.

Similar with modal verb *can*, *could* is mostly used in spoken text. Seeing its functions which is less than the function of verb *can*, it is relevant that the occurance frequency of *could* (1,721) is less than *can* (3,089).

Will

Used for talking about or predicting the future; used for showing that somebody is willing to do something; used for asking somebody to do something; used for ordering somebody to do something; used for stating what you think is probably true; used for stating what is generally true; used for stating what is true or possible in a particular case; used for talking about habits.

The modal verb *will* is the mostly used in written text of all over the other modal verbs. The difference in frequency between written text (2,587) and spoken text (1,557) is quite noticeable.

Might

Used as the past tense of *may* when reporting what somebody has said; used when showing that something is or was possible; used to make a polite suggestion; used to ask permission politely; used to ask for information; used to show that you are annoyed about something that somebody could do or could have done; used to say that you are not surprised by something; used to emphasize that an important point has been made.

Come near to *shall*, the modal verb *might* is not used very much in both spoken and written text. It is used more dominantly in spoken text than written text. The comparison of occurance frequency is also not very visible between spoken text (618) and written text (400).

Should

Used to show what is right, appropriate, etc., especially when criticizing somebody's actions; used for giving or asking for advice; used to say that you expect something is true or will happen; used to say that something that was expected has not happened; used after *I* or *we* instead of *would* for describing what you would do if something else happened first; used to refer to a possible event or situation; used after *that* when something is suggested or arranged; used after *that* after many adjectives that describe feelings; used with *I* and *we* in polite requests; used with *I* and *we* to give opinions that you are not certain about; used for expressing strong agreement; used to refuse something or to show that you are annoyed at a request; used to express surprise about an event or a situationused to refuse something or to show that you are annoyed at a request; used to express surprise about an event or a situation; used to tell somebody that something would amuse or surprise them if they saw or experienced it.

As can be seen above, there are a lot of functions of this modal verb *should* which is elaborated by oxford dictionary, but its use in the text is still below the use of the other four modal verbs (*can, would, will,* and *could*). It is used more dominantly in written text than spoken text. The comparison of occurance frequency is also not very visible between written text (1,187) and spoken text (1,037).

Would

Used as the past form of *will* when reporting what somebody has said or thought; used for talking about the result of an event that you imagine; used for describing a possible action or event that did not in fact happen, because something else did not happen first; used for saying why somebody does something; used for saying what you want to happen; used to show that somebody/something was not willing or refused to do something; used to ask somebody politely to do something; used in polite offers or invitations; used to say what you like, love, hate, etc.; used to give opinions that you are not certain about; used to give advice; used for talking about things that often happened in the past; used for talking about behaviour that you think is typical; used to express a strong wish.

In spoken text, the frequency of the modal verb *would* is the second highest after *can*, and also the second highest after *will* in written language. *Would* is used more dominantly in spoken text than written text.

From the explanations above, it can be concluded that the spoken text uses more modal auxiliary verbs than the written text. The tendencies of spoken text are to express permission, possibility, ability (relected by modal verbs *can, could,* and *might*), volition and prediction (reflected by modal verbs *shall* and *would*). Meanwhile the written text tends to express obligation, necessity (reflected by modal verbs *must* and *should*), permission, possinility (reflected by modal verb *may*), volition and prediction (reflected by modal verb *will*).

4.2 Personal pronouns which followed by modal verbs

Spoken Text

After searching the structure of the clauses consisting of personal pronoun + modal verb through CQPWeb, as presented in Figure 3, suggests that the personal pronoun that appears the most before a modal verb is *I,* followed by *you, we, it, they, he*, and the least personal pronoun is *she*. For further details of total frequency data can be seen from Table 3.

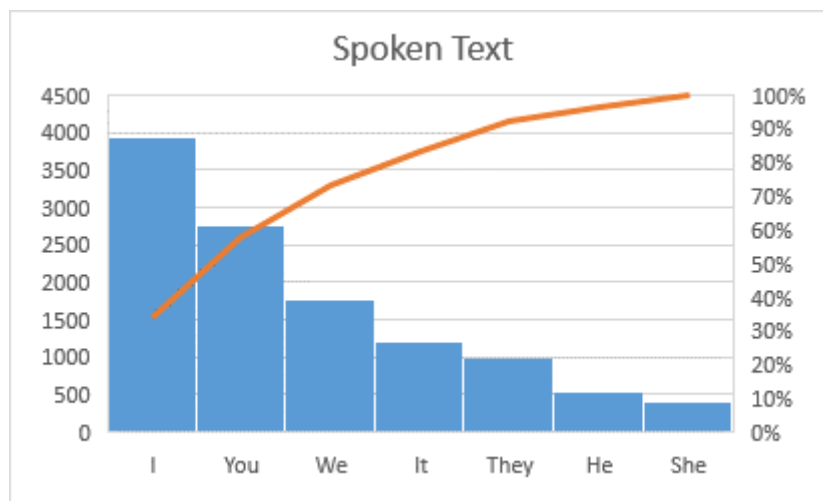**Figure 3**

Table 3

SPOKEN TEXT		
Personal Pronoun	**Frequency**	**Token**
I	3935	4487
We	1770	2019
You	2761	3149
He	538	614
She	394	450
They	989	1128
It	1196	1364

From explanation above it can be concluded that the spoken text tends to refer to first person singular subjective personal pronoun *I*, in order to express the speaker's feeling or belief againts something. The second most frequent pronoun which appears is *you*, to refer to the reader or listener or participant when the text is delivered. The caluses can be suggestion, permission, order, from the speaker to listener or participant.

Written Text

Here is the finding of the clauses consisting of personal pronoun + modal verb in written text through the same steps as previously described. For the personal pronoun that appears the most and the least is similar with the finding before, they are *I* and *she*. Still, there is a difference among the order of the pronouns between them. The second highest number of personal pronoun that appears in written text is *it*, followed by *you, we, they, he,* and the last is *she* (see Figure 4). For further details of total frequency data can be seen from Table 4.
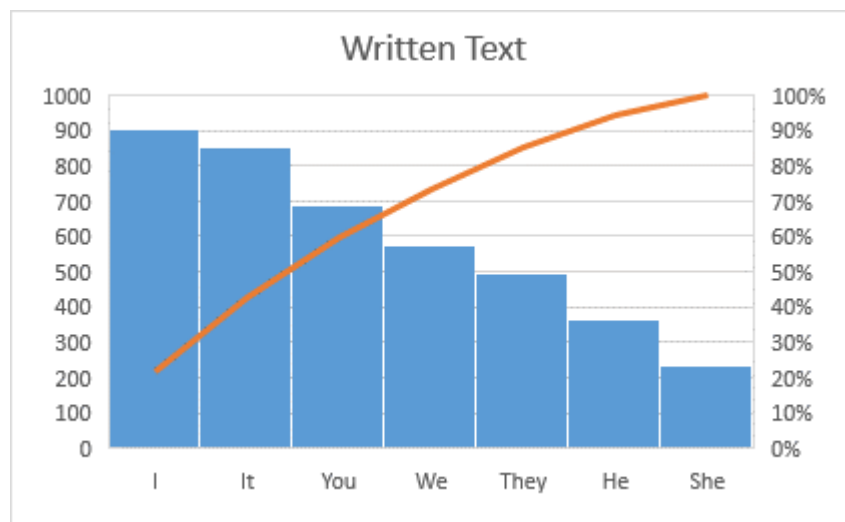

Figure 4

Table 4

WRITTEN TEXT		
Personal Pronoun	Frequency	Token
I	903	1052
We	573	668
You	689	803
He	363	423
She	236	275
They	496	578
It	850	990

As can be seen from the data above, the most personal pronoun appears in written text is also *I*, to refer to the writer's feeling or belief againts something. Differ from previous finding, the second most frequent pronoun which appears in written text is *it*. Pronoun *it* can refer to many

things, moreover in written text. It can be concrete noun or even abstract noun. It is undeniable that *it* occupies the second most frequency position after *I*.

5 Conclusion

The researchers conclude that there are different trends between spoken text and written text in terms of the use of modal auxiliary verbs. The tendencies of spoken text are mostly to express permission, possibility, and ability. Meanwhile the written text mostly tends to express obligation, and necessity. Beside there are several other trends that have been mentioned previously. For the second research question about personal pronouns which occur before modal verbs, there are similarities and differences between spoken text and written text. The similar finding is the personal pronoun that appears the most and the least are *I* and *she*. The different finding is the second most frequent pronoun which appears in spoken text is *you*, meanwhile in written text is *it*.

REFERENCES

[1] Crawford, Csomay, E., & J, W. (2017). Review: Crawford and Csomay (2016). Doing Corpus Linguistics. *Corpora*, *12*, 307–309. https://doi.org/10.3366/cor.2017.0120

[2] Leech, G. (2004). *Meaning and the English Verb—3rd Edition—Geoffrey N. Leech—Routl*. https://www.routledge.com/Meaning-and-the-English-Verb/Leech/p/book/9780582784574

[3] Lindquist, H. (2009). *Corpus Linguistics and the Description of English*. Edinburgh University Press.

[4] Lukitawati Sujatna, M., Tuckyta Sari Sujatna, E., & Pamungkas, K. (2019). Exploring The Use Of Modal Auxiliary Verbs In Corpus Of Contemporary Of American English (COCA). *Sosiohumaniora*, *21*(2), Article 2. https://doi.org/10.24198/sosiohumaniora.v21i2.19970

[5] Nugraha, D. N. S. (2019). *Double Modal Verbs (Have to) and Verbs in English Modality on Corpus Linguistics in the Context of Industrial Revolution 4.0: Syntactic and Semantic Studies*. *4*(5).

[6] Oktavianti, I. N. (2019). *A Corpus-Based Analysis Of English Core Modal Verbs And Their Counterparts In Indonesian*. *8*(12).

[7] Portner, P. (2009). *Modality, Oxford University Press, 2009*. http://roozbehhormozi.persiangig.com/document/Modality,%20PAUL%20PORTNER,%20Oxford%20University%20Press,%202009.pdf/dl

[8] Rizvic-Eminovic, E., & Šukalić, Đ. (2019). CORPUS-BASED STUDY OF THE Modal Verbs In The Spoken And Academic Genres Of The Corpus Of Contemporary American English. *Zbornik Radova 18*. https://doi.org/10.51728/issn.2637-1480.2019.17.351

[9] Stefanowitsch, A. (2020). *Corpus linguistics: A guide to the methodology*. Zenodo. https://doi.org/10.5281/ZENODO.3735822

[10] Wehmeier, S., & Hornby, A. S. (2005). *Oxford Advanced Learner's Dictionary: Of Current English*. Oxford University Press.

[11] Winiharti, M. (2012). The Difference Between Modal Verbs in Deontic and Epistemic Modality. *Humaniora*, *3*(2), Article 2. https://doi.org/10.21512/humaniora.v3i2.3396