JALL (Journal of Applied Linguistics and Literacy)
English Education Program
Faculty of Teacher Training and Education
Galuh University
ISSN 2598 - 8530
Jl. R.E. Martadinata No. 150 Ciamis 46251
jall@unigal.ac.id
https://jurnal.unigal.ac.id/index.php/jall/index
JALL (Journal of Applied Linguistics and Literacy), ISSN 2598-8530, September, Vol. 7 No. 2, 2023
Received: June 22nd, 2023. Accepted September 18th, 2023. Published September 21st, 2023

MODAL AUXILIARIES AS EPISTEMIC DEVICES IN MARKING SCIENTIFIC RESEARCHERS' UNCERTAINTY ON COVID-19

Nur Azwin Zulkarnain
Ridwan Wahid*
Jariah Mohd Jan

Universiti Malaya, Malaysia

ridwanwahid@um.edu.my

ABSTRACT

The early phase of the COVID-19 pandemic caused a lot of pressure on scientists due to the novel nature of the coronavirus. As experts in the field, they were expected to produce only reliable information. Owing to the limited data available at the time, there were many uncertainties surrounding the virus. However, studies that looked into how the uncertainties were navigated are scarce. This corpus-based study investigates this issue using the system value of modal operators by Halliday and Matthiessen (2014), along with Dong et al.'s (2020) classification of COVID-19 research themes to explore levels of scientific researchers' certainty in presenting information about coronavirus. Specifically, their choices of modal auxiliaries as epistemic devices are analyzed. Results demonstrate that researchers mainly conduct studies on epidemiology with the lowest degree of certainty by utilizing models such as may, could, and might. Furthermore, while some of the propositions expressed do display researchers' assumptions of possibilities, they, however, are presented with insufficient through their use of epistemic devices and contribute to a better understanding of their intentions in conveying vital information.

Keywords: *Modal auxiliaries; Epistemic devices; COVID-19 research; Uncertainty*

INTRODUCTION

During the early breakout of COVID-19, numerous cases of infections were inadvertently left undetected due to the limited capacity of scientists to carry out tests, as well as the absence of recognisable symptoms among those who were infected (Brown & Walensky, 2020). As the number of infections rose, the uncertainty of vital information about the virus and its variants led to questions regarding how long and to what extent it would affect the health of the community (ibid.). With the constant worry and fear about the spread of COVID-19 among community members, scientists did their part and conducted numerous studies to obtain a better understanding of the virus' life-threatening nature. The research focus on COVID-19 has been consistent since early 2020 (Iacobucci, 2020), matching rising anxiety levels among society members (Amsalem et al., 2021). It was therefore vital for scientific researchers to address COVID-19 with appropriate utilisation of epistemic devices in order to indicate reliability.

According to Correia (2020), the novel coronavirus has negatively affected society in so many ways that the presentation of information regarding it is considered important not just to the public in general, but to the scientific community as well. Further, physicians investigating COVID-19 believe that the public depends on their expertise and thus, it becomes imperative that the information they present be reliable despite the possibility of bearing a high degree of uncertainty (Orso et al., 2020). Due to the pressure of being depended on, these experts' quality of work was often questioned. For instance, the trust in COVID-19 literature was claimed to be flawed by the lack of competence in quality research publications and findings (Teixeira da Silva, 2020). Claims such as this raise questions among readers and also those within the scientific community. Ultimately, the doubts can also affect the entire population (Teixeira da Silva, 2021).

Tran et al. (2020) drew attention to scientists' higher interest in clinical investigations that encompassed discussions like emergency care and management, as well as studies of previously detected coronaviruses. Topics that were least discussed on the other hand, were more inclined towards the impacts of the virus and tests for efficacies. Studies investigating how certainty is expressed by these experts in addressing these topics were even more limited and hence, the subject of the present research.

Following Hyland (1996), the semantic domain of epistemic modality correlates with certainty of knowledge and is realised by epistemic devices. These devices, which primarily include the modal auxiliaries, are different types of expressions that convey the degree of one's certainty in the asserted message. As pointed out by Halliday and Matthiessen (2014), in systemic functional grammar, modality represents the intermediate degrees that lie between the 'yes' and 'no' continuum in the interpersonal function of language. In addition, Adegbola (2019) demonstrated that epistemic modality markers do not merely convey factuality of information but also act as an ideological tool to show a speaker's or writer's point of view. Besides, the relationship between modality and point of view is important to meaning (Simpson, 1993, p. 42). On this account, modal auxiliaries as epistemic devices are recognised to exhibit different levels of "strength" that correlate with one's certainty and the truth probability of a proposition (e.g., *will* – highest, *could* – lowest) (Winiharti, 2012).

The role of epistemic modality as a means to provide textual evidence in discourse is undoubtedly crucial. Hyland (2005a) argued that epistemic devices that increase or decrease the strength of an argument are highly associated with three important

speaker/writer factors: (i) willingness to entertain the prospect of alternatives, (ii) level of commitment to the propositional content, and (iii) respect for addresses' participation in the argument. It should also be noted that use of epistemic modality in research articles differs greatly from one field to another. In the scientific context for instance, Martín-Martín (2008) argued that rhetorical epistemicity is often related to the strength of the scientific assertions which, in turn, have an effect on the community's acceptance. Given this background, the study aims to investigate the use of modal auxiliaries as epistemic devices in demonstrating scientific researchers' (un)certainty in various research foci pertaining to the novel COVID-19.

LITERATURE REVIEW

COVID-19 Uncertainties and Research Focal Points

Scientific information on the COVID-19 pandemic is regarded to be indispensable considering the severity of the health threats imposed on populations of the world. Yeo-Teh and Tang (2020) pointed out the crucial role of epidemiologists as experts in the field to pursue nothing but only reliable findings to avoid wasting effort, time, as well as resources. Besides, information that is known of its source, which is input from epidemiologists, is expected to be free from any dubiousness (Orso et al., 2020). This demonstrates the heavy responsibility shouldered by these experts as uncertainties from them could aggravate the consequences of the pandemic (Koffman et al., 2020).

In addition, science as part of policymaking has a strong influence on public trust in its role (Kreps and Kriner, 2020). Thus, while scientific writers do acknowledge uncertainty in their work, they have to consider whether it actually boosts public confidence and support for science-based policymaking or increase cynicism towards science due to excessive and irrational uncertainty (ibid.).

As the research on COVID-19 expanded, various topics were explored. A recent study by Tran et al. (2020) investigated the research foci on COVID-19 and identified a variety of approaches and measures taken by countries with different levels of income and transmission to fight the virus. A textual analysis and the Latent Dirichlet allocation (LDA) to identify topic modelling were carried out on research articles published between December 2019 to April 2020 and it was discovered that topics on epidemiology were most commonly addressed.

Another study that utilised LDA for the purpose of topic modelling on COVID-19 and related coronavirus research abstracts is Dong et al. (2020) who identified the frequent and infrequent research areas on the virus. Dong et al. (2020) suggested the utility of studying characteristics of research publications on coronavirus across many years to identify the research gaps that could offer useful information for the prevention of future outbreaks. By opting for a topic modelling analysis on article abstracts of a corpus obtained from the COVID-19 Open Research Dataset (CORD-19), eight high priority scientific research themes were discovered namely, *clinical characterisation*, *pathogenesis research*, *therapeutics research*, *epidemiological study*, *virus transmission*, *vaccines research*, *virus diagnostics*, and *viral genomics*. These themes were classified in the order of their prevalence and based on the similarity of the semantic content of the fifty most frequent words that occurred in the corpus. For instance, the semantic categorisation of *clinical characterisation* was based on words such as 'infection', 'disease', 'cause', 'severe', etc. Further, Dong et al. (2020) also noted how such a category was strongly related in all the other seven topics, signifying the vital role of clinical characterisation in studies about the coronavirus. According to Li et al. (2021), scientists began to focus on clinical characteristics of the virus only after the number of confirmed cases increased drastically as the pandemic progressed.

Modal Auxiliaries as Epistemic Devices in Research Articles

Halliday and Matthiessen (2014) presented a modal system that represents various degrees of realisation in communication, which is related to epistemic modality. This refers to the probability (i.e., certain, probable and possible) of a proposition to be factual and true. Moreover, epistemic modality use in scientific discourse is said to be a rhetorical tool to reflect the degree of commitment of the writer to the truth, as well as to minimize criticisms from experts in the field (Martín-Martín, 2001). All these studies point to the importance of such linguistic features in the scientific writing convention.

Epistemic devices that can be classified as hedges, approximators, and booster are useful in a way that they are also commonly used to convince and influence readers in the field (Yang et al., 2015; Rozumko, 2017). Flowerdew and Peacock (2001) argued that though the concept of epistemicity is similar across languages, the different key elements in stance-making are the determinant that makes rhetorical patterns distinctive from one discourse to another despite containing the same subject matter. As such, past studies noted

on the different epistemic modality preferences by academic writers that were reported to be influenced and were restricted by the discourse norms and rhetorical styles of their respective disciplines (see, e.g., Rozumko, 2017; Akbas & Hardman, 2018). On this account, Hyland (2005a) regards epistemic devices as researchers' way of expressing their judgements on the status of propositions while having them as engagement markers that connect them with the addressees to either only draw their attention to the content of the argument with utmost certainty, or consider them as essential members or participants of the discourse to elicit insights on their interpretations.

As far as hedges and boosters as devices of epistemic modality are concerned, it was reported that the former appear to be more central to epistemicity, especially in academic and scientific discourses (Akbas & Hardman, 2018). The relationship between hedges and epistemic modality is established due to the purpose of the former that is shaped based on mental attitude with the intention of vagueness, which is drawn from modals (Salager-Meyer, 1994). As such vagueness does represent the writer's commitment to the proposition, Hyland (1996) argues hedges are essential in scientific discourse to maximise reliability and precision of the writer's certainty and to show clarity with regard to a state of knowledge. This fact assists readers in increasing the possibility for them to interpret the proposition exactly in the way the writer depicts it. Furthermore, Martín-Martín (2008) highlighted the practicality of both hedges and boosters as crucial rhetorical elements that are widely practised by experts in conveying health-related information. This fact signifies the importance of these devices that represent the writers' sensitivity towards the understandings of their social communities, as well as contributing knowledge to the public and their disciplinary societies (Hyland, 2005a).

Halliday and Matthiessen (2014) termed epistemic devices as modal operators and distinguished three levels of how these devices could be articulated where the system value also incorporates the negation of propositions and modalities (see Table 1).

Table 1. System value of modal operators as epistemic devices (Halliday & Matthiessen, 2014)

Degree of Certainty	Epistemic Modality Value	Modals
High	Certain	must, ought to, need, has/had to, mustn't, oughtn't to, can't, couldn't, may not, mightn't, hasn't/hadn't to
Median	Probable	will, would, should, won't, wouldn't, shouldn't
Low	Possible	can, may, could, might (dare), needn't, doesn't/ didn't need to, have to

The negated properties of these modals were classified by Halliday and Matthiessen (2014) due to the existence of polarity as an important feature in language finiteness. Modality is regarded as a mechanism to express the degree of probability (of a proposition) that occurs within the semantic space in which an exchange takes place between the interlocutors (ibid.). This explicates how modals of different tenses could appear in either a positive form or a negated one according to their respective epistemic modality values. Despite that, Ngula (2017) affirmed that the modal *can* (including its negated form) is argued to express little semantic properties of epistemicity, claiming it is best for it to be excluded from the list of epistemic devices.

As in Hengeveld and Mackenzie (2008), different levels of certainty and commitments are regarded as the results of modifiers of the speaker/writer's attitudes,—which are related to the domain of subjective epistemic modality. Hence, scientific assumptions and hypotheses of the ever-changing natural phenomena are believed to have an influence on scientific researchers' evaluative linguistic choices that mark their degrees of epistemic certainty (Hyland, 2005b).

METHODOLOGY

Following the aim of the research which is to elucidate the semantics of modal auxiliaries as epistemic devices in denoting scientific researchers' certainty in different COVID-19 research foci, the study adopts a corpus-based approach (see also Yang et al., 2015; Rozumko, 2017; Akbas & Hardman, 2018). Based on Halliday and Matthiessen's (2014) list of modal operators that convey the concept of probability, the study employs the predetermined modal auxiliaries to show degrees of certainty and epistemic modality values. However, the modal *can* and its negated counterparts are excluded from analysis due to their lack of epistemicity. Apart from that, the study adopts Dong et al.'s (2020) eight research themes in order to identify COVID-19 topics which were addressed with (un)certainty. The following figure represents the overall research design of the study, starting with the compilation of research articles to build a corpus, and ends with the identification of discernible research foci in accordance with researchers' certainty levels.

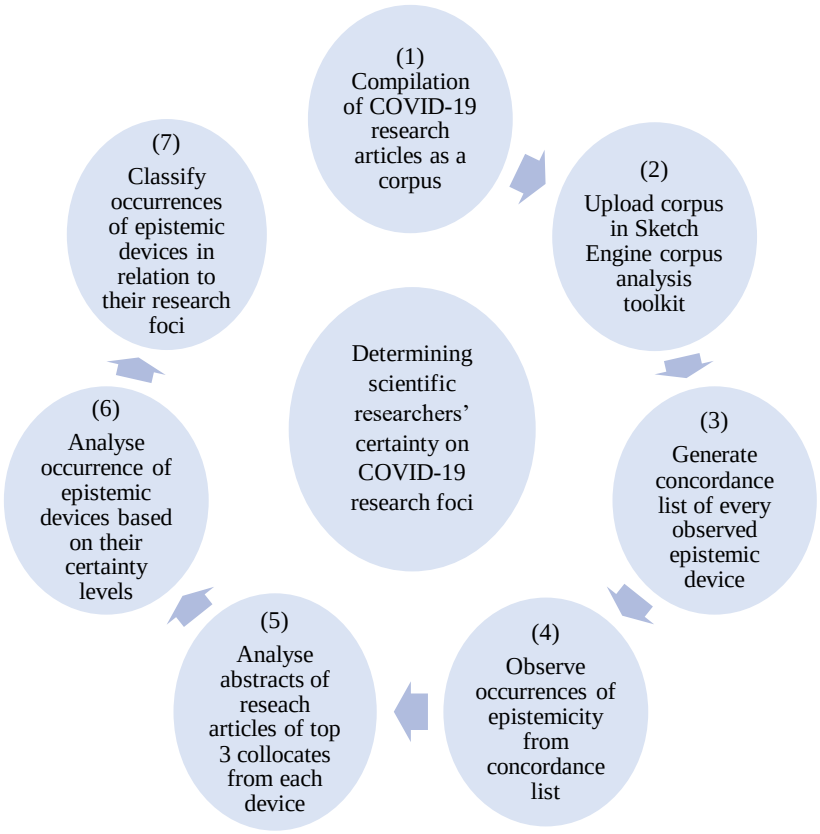


Figure 1. Overall view of study's research design

As can be seen in Figure 1, the study employs a detailed research design in order to correlate the semantics of epistemic modality as a means of expressing researchers' certainty with the reported various research foci on COVID-19 via a corpus-based analysis. The analysis enables the discernment of propositions in which the epistemic devices occur in the corpus data in relation to their epistemic modality values, and directly identifies the research foci they represent. The epistemic modality values follow Halliday and Matthiessen's (2014) system of modality operators that encompass the value of *possible*, *probable* and *certain*. Subsequently, these values are associated with Dong et al.'s (2020) 8 COVID-19 research themes that emerged during the early phase of the pandemic (see Figure 2).

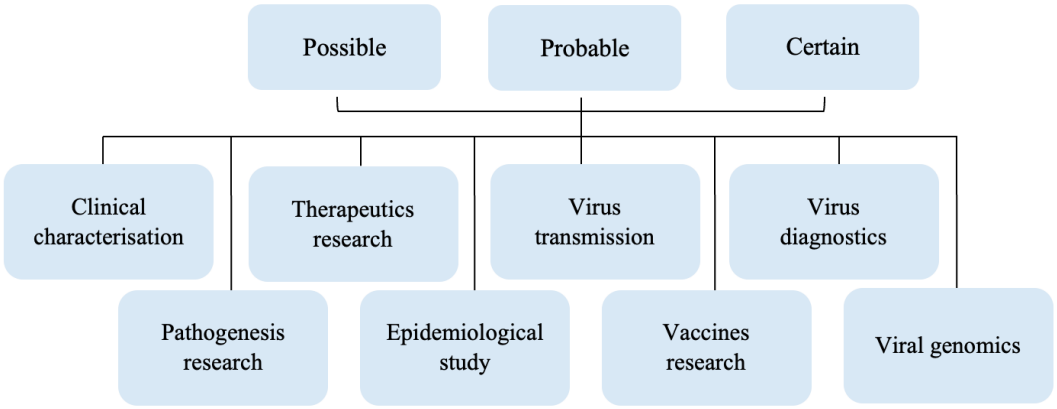


Figure 2. Framework of study

Research Data

The research data utilised in the study is a small corpus that was built based on a compilation of scientific research articles about COVID-19. It consists of 20 research articles that were results of the keyword search "COVID-19 pandemic" and "uncertainty" on Google scholar (https://scholar.google.com), and they were randomly selected to maintain a fair distribution. The small corpus contains 100,401 words and 138,972 tokens altogether. It was then uploaded on the Sketch Engine website, an online corpus analysis toolkit.

Data Analysis

It is important to note that the study employs a corpus-based semantic analysis where the procedure includes both quantitative and qualitative methods. The quantitative analysis deals with ascertaining the frequency of (un)certainty expressions through the use of epistemic devices by researchers. The qualitative method involves the classification of each research articles according to COVID-19 research themes, as well as analysing the occurrences of the epistemic devices in the context of those themes.

To illustrate, the 25 epistemic devices that comprise epistemic modal auxiliaries were analysed in terms of how they occurred in context. This was done by utilising the concordance tool, though only epistemic devices that occurred more than 10 times in the data were observed in the study. This is because only abstracts of research articles from top three collocates of each epistemic device were analysed to identify the themes of the research. Further, each analysed abstract was based on a sample chosen at random from each of those three collocates. Thus, the epistemic devices that did not occur at least 10 times in the corpus data were disregarded to avoid the possibility of interference in finding the relevant top three collocations for those devices. For this reason, the study performed the

analysis on collocations that were ranked based on the MI-score calculated within a span of three collocates to the left and right of the search word (-3, +3). MI-score is effective in finding word combinations that strongly co-occur, rather than word pairs that were randomly distributed (Hunston, 2002). These collocates were further semantically examined based on the context that they occurred in. The most relevant occurrences of the collocates were chosen to assess their categories of research foci, which was accomplished by analysing the content of the article's abstract. This is consistent with one of the methods undertaken by Dong et al. (2020) as a means for topic modelling.

To validate this analysis procedure, a coder with a professional scientific background was recruited to verify the coding of coronavirus research foci. Following it, an interrater reliability test was performed. The test results showed a similarity index of 91%.

FINDINGS AND DISCUSSION

Frequency of Modal Auxiliaries as Epistemic Devices

Based on the analysis, a total of 842 occurrences (8.386 per thousand words (ptw)) of modal auxiliaries were found (see Table 2). In terms of epistemic modality values, most writers/researchers were found to express a possible value (4.024 ptw), which represents the lowest value of certainty. This is followed by the median degree of certainty, the probable value, with 3.197 ptw. Next, the least preferred use of epistemic devices were those that express the certain epistemic modality value (1.165 ptw), indicating the highest degree of certainty. Based on these frequencies, it can be said that the pattern of modal auxiliaries as epistemic devices used by these experts was more inclined towards expressions denoting less than full certainty.

Table 2. Observed frequencies of epistemic devices

Epistemic Modality Value	Number of Occurrence	Frequency ptw
Possible	404	4.024
Probable	321	3.197
Certain	117	1.165
Total	842	8.386

Correspondingly, the following Table 3 shows the frequency of the 25 observed epistemic devices.

Table 3. Frequencies of occurrences for each modal auxiliary

Epistemic Modality Value	Modal	Frequency of Occurrence	Frequency ptw
Possible	may	272	2.709
	could	67	0.667

	might	55	0.548
	have to	8	0.080
	need not	1	0.010
	does not need to	1	0.010
	did not need to	0	0.000
Probable	will	116	1.155
	should	114	1.135
	would	76	0.757
	should not	7	0.070
	will not	5	0.050
	would not	3	0.030
Certain	need	71	0.707
	must	19	0.189
	may not	13	0.129
	had to	6	0.060
	ought to	4	0.040
	might not	2	0.020
	must not	2	0.020
	could not	0	0.000
	has to	0	0.000
	ought not to	0	0.000
	has not to	0	0.000
	had not to	0	0.000
Total		842	8.836

The analysis suggests that the modal *may* was most frequently applied in the corpus data with 2.709 ptw. Specifically, 32.3% of the total was the use of *may* as a means to express researchers' state of least certainty. However, the second most used modal was the modal *will* (1.155 ptw), followed by *should* (1.135 ptw), where both represent the probable value. Apart from that, it is also evident that not all of the epistemic devices were utilised by the researchers. These include *did not need to, could not, has to, ought not to, has not to,* and *had not to.* The majority of them are negated modals and were not identified in the corpus. Moreover, most of these absent modals mark a high degree of certainty, which explains their non-occurrence.

Scientific Researchers' Certainty on COVID-19 Research Foci

Using the adopted criteria of collocations, only 24 collocates of the occurrences of epistemic modality qualified and were subsequently categorised according to their respective research themes. As Dong et al.'s (2020) categories of coronavirus research focus consisy of matters pertaining to both the novel COVID-19 and past coronavirus, the application of the framework in the current study centring only on COVID-19 ascertained the influence of the pandemic on the direction of the research at the time. This is due to the theme *epidemiological study* that was perceptible to be scientific researchers' common research focus to be discussed, particularly in the state of being less certain (see Table 4).

Table 4. Frequencies of Scientific Researchers' Certainty Based on COVID-19 Research Foci

COVID-19 Research Focus	Possible	Probable	Certain	Total
Epidemiological study	4	3	2	8
Clinical characterisation	2	3	1	7
Therapeutics research	1	2	2	5
Miscellaneous	1	0	1	2
Vaccines research	0	0	0	0
Pathogenesis research	0	0	0	0
Virus transmission	0	0	0	0
Virus diagnostics	0	0	0	0
Viral genomics	0	0	0	0
Clinical characterisation + Epidemiological study	1	1	0	2
Total	9	9	6	24

As shown in Table 4, topics on epidemic situations and diseases control accumulate the highest total of occurrences with epistemic devices (N=8) where most of them are addressed with either a possible or probable value. Scientific researchers were found to generally emphasise the situation of the pandemic and its effects on nation leaders, healthcare workers, and the public in the said research theme.

By way of explanation, the top 3 collocates that cooccur with the modal *might*, which signifies researchers' lowest level of certainty, were found in a research article that was categorised as a study on epidemiology (see Figure 3).

	Word	Cooccurrences	Candidates	MI ↓
1	different	3	65	6.87
2	be	20	436	6.86
3	one	4	93	6.76

Figure 3. Collocates of the modal *might*

The cooccurrence of *might* with the collocate *different* in the data has the highest MI-score among other collocates. Based on the results of the analysis, the use of the modal with *different* in the data demonstrates the intention of researchers to make assumptions on a situation that may contradict the actual state. As such, propositions that are regarded as mere possibilities were asserted in a manner that avoided absolute certainty. This may have to do with the fact that the COVID-19 was still novel at the time and a lot of the information could not be confirmed. Some examples are provided in Figure 4.

	Details	Left context	KWIC	Right context
1	doc#8	ɜt to public trust when multiple reversals aggregate.</s><s>These scenarios	might	include different types of reversals and consider the erosion of trust when th
2	doc#19	's><s>The cumulative evidence here suggests that very different strategies	might	be called for in varying cultural contexts in the fight against COVID-19.</s><
3	doc#19	ɔublic health.</s><s>One issue with polarization during a pandemic is that it	might	lead different segments of the population to arrive at different conclusions a

Figure 4. Sample concordance lines for *might + different*

Lines 1, 2 and 3 in Figure 4 exhibit how the phrase *might + different* occurs in the corpus data. As can be seen, all three samples suggest the state of possibility in relation to types of reversals, strategies and segments of populations involved in the COVID-19 pandemic. These propositions were asserted with an intention that they could provide an alternative view or perception of what was being discussed. However, the presentation of such propositions was careful and not forced on the readers.

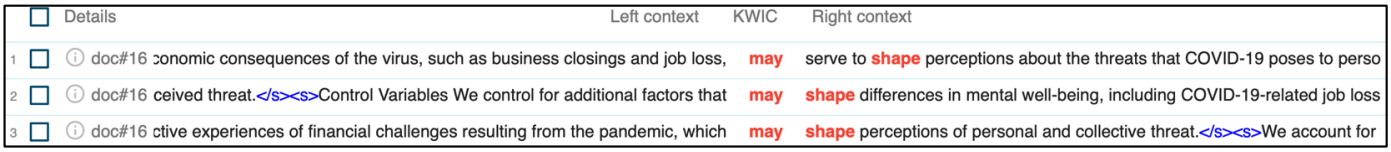
Excerpt 1

Figure 4, Line 1	**Example in Context**
might + different	"Our results suggest that the presentation of the scientific uncertainty in the media and how other actors seek to politicize it affects both immediate public support for using COVID-19 models to guide policymaking and public attitudes toward science more generally. The lasting implications of the latter could be particularly significant. Our experiments considered COVID-19 model projections and reversals, but scholars might also consider the potential cost to public trust when multiple reversals aggregate. These scenarios **might** include different types of reversals and consider the erosion of trust when the CDC reverses guidance on masks and projected fatalities."

For instance, the context in Line 1 as shown in Excerpt 1 presents the scientific researchers' discussion on the many conditions in which scientific uncertainty could have an effect on public trust in science and science-based policies during the pandemic. They also discuss the issue of projections and reversals although not all reversals are highlighted in detail. The concern about the risk of a declining public trust is further elaborated through the possibility of various different reversals that may affect the public significantly. These experts establish the fact that they may not have the knowledge of the consequences of all potential reversals, but still flag the possibility for them to occur. Thus, the use of *might* as an epistemic device in the proposition serves as a means to tell the scientific community there is no certainty that can be established about it.

Additionally, the most frequent modal that expresses uncertainty, which is *may*, correlates primarily with research pertaining to finding clinical characteristics of the virus. This type of research was regarded as essential in the process of understanding the virus' lethal nature as according to Dong et al. (2020), almost all the other research foci were

strongly related to the said topic. Considering the fact that COVID-19 was unfamiliar during its early detection, information on the virus' symptoms and effects were of uttermost importance to assist the research community with regard to elucidating other aspects of the virus as well as the pandemic, which explains the prevalence of studies on epidemiology. To explicate, the third collocate for the modal *may*, *shape* is presented in Figure 5.



	Details	Left context	KWIC	Right context
1	doc#16	conomic consequences of the virus, such as business closings and job loss,	may	serve to shape perceptions about the threats that COVID-19 poses to perso
2	doc#16	ceived threat.</s><s>Control Variables We control for additional factors that	may	shape differences in mental well-being, including COVID-19-related job loss
3	doc#16	ctive experiences of financial challenges resulting from the pandemic, which	may	shape perceptions of personal and collective threat.</s><s>We account for

Figure 5. Sample concordance lines for *may* + *shape*

With an MI-score of 7.77, the cooccurrence of *may* with *shape* in the corpus data is essentially used by researchers to discuss the possibility of a deduction made by the researchers in relation to the effect that the virus has on the mental wellbeing of the public after being exposed to constant updates about the virus' progress.

In line 1 for instance, the researchers argue that the situation at the time is triggering anxiety among the public as the vaccine is yet to be discovered and questions concerning the severity of the virus are dominant.

Excerpt 2

Figure 5, Line 1	**Example in Context**
may + shape	"We suggest that COVID-19 news media consumption, especially in the absence of a solution to the virus, such as a vaccine, may amplify the perceived threats the virus presents to personal and population health and well-being. The repeated viewing of information about the number of new infections, deaths, and strained medical facilities, as well as the economic consequences of the virus, such as business closings and job loss, **may** serve to shape perceptions about the threats that COVID-19 poses to personal and public wellbeing. This is particularly true with regard to health and finances."

The approach taken by the researchers in conveying their verdict about the possibility of the public's negative perspectives on the virus is initially discussed by pointing out that the pandemic progressively affects lives, the amount of medical facilities is shrinking, as well as the economy is deteriorating. Based on the evident impact of the virus, the researchers conclude that it could therefore contribute to the decline of the public's mental wellbeing as too much negativity is evoked in the process. Despite that, the presentation of the proposition is not directly established by the experts to imply that such facts are absolute. Instead it is carefully asserted as a suggestion using the modal *may*. This is due to other

possibilities that might also contribute to the public's perceptions which have not been discussed.

Moreover, in terms of showing a probable value of certainty, findings show that researchers are more inclined towards conducting studies pertaining to clinical characterisation. These research articles are written in a manner where more certainty is presented, although maximum certainty is still not guaranteed. As previously mentioned, Dong et al. (2020) reported that this research focus is vital as its development greatly contributes to the conducting of other foci of research, a fact which elucidates the importance for its findings to be eminently reliable. Hence in order for their findings to appear trustworthy enough for other researchers to make reference to, a higher likelihood of certainty should be demonstrated. To give an instance, the most frequent modal that conveys a median level of certainty is the modal *will* with 1.155 ptw. Accordingly, Figure 6 shows the collocates that cooccur with *will*.

	Word	Cooccurrences	Candidates	MI ↓
1	definitely	3	6	9.23
2	continue	5	12	8.96
3	say	4	13	8.53
4	require	3	12	8.23
5	influential	3	13	8.11

Figure 6. Collocates of the modal *will*

Based on the analysis of each occurrence, it was found that the collocates *definitely* and *say* do not represent propositions from the perspective of the researchers. Instead, it is simply a report of their samples' opinions on the need for them to get vaccinated. Due to this, these collocates are disregarded. The subsequent collocate after *definitely*, i.e., *continue*, also strongly correlates with *will* where it identifies the state of researchers' certainty with regard to a kind of prediction (see Figure 7).

	Details	Left context	KWIC	Right context
1	doc#6	vill look to extend research contracts and have "an open mind" on whether it	will	continue to fund additional research costs beyond existing schedules.</s><
2	doc#8	ıealth guidelines.</s><s>Given the novelty of the COVID-19 virus, scientists	will	continue to accrue more data, identify missing variables, and learn more ab
3	doc#10	he immediate impacts of the pandemic at a single time-point, circumstances	will	continue to evolve, and there will likely be other notable impacts to science.
4	doc#15	andated social isolation in many regions, in a climate of fear, social isolation	will	continue to persist after the risk of infection has substantially declined.</s><
5	doc#16	ffects tend to wane, especially among those not directly involved, COVID-19	will	continue to produce uncertainty and threat into the foreseeable future.</s><

Figure 7. Sample concordance lines for *will + continue*

As established in Figure 7, the cooccurrence of the modal *will + continue* in the data mainly suggests the level of willingness of scientific researchers in making predictions about situations on the COVID-19 virus. Due to the immense inconsistency and rapid mutation of the virus, it seems like researchers are unable to reach a consensus about the impacts of the virus in relation to various concerns. This includes matters about research fundings, scientists continued effort in acquiring data, public's wellbeing, and the vast unforeseeable threats of the virus. All of these concerns relate to the epidemiological considerations that are the outcomes of understanding the characteristics and effects of COVID-19. In Line 4 for example, the researchers raise the issue of fear that the public has towards the virus and how it could have an effect on their mental health.

Excerpt 3

Figure 7, Line 4	**Example in Context**
will + continue	"Fear in these cases serves to isolate individuals (de Rivera, 1992). In a similar way, the uncertainty caused by the COVID-19 pandemic may lead to a climate of fear within the United States and across the globe. Although governments have mandated social isolation in many regions, in a climate of fear, social isolation **will** continue to persist after the risk of infection has substantially declined."

While most studies concentrate on the symptoms of being infected by COVID-19, this research emphasises the fact that patients are not only physically affected by the virus, but also mentally, due to fear. The decline in mental health among these individuals may have been caused by traumas or phobias stemming from the likelihood of them being infected again or spreading the virus to their loved ones. As such, the implementation of social isolation is regarded as necessary as an approach to confront the issue. The researchers continue by predicting the indispensability of the approach until the number of those being infected decreases. Despite not being able to provide a definite prognosis due to the uncertainties, the modal *will* represents the researchers' prediction, as well as judgement on what is regarded as essential in flattening the curve for COVID-19. It serves as a mechanism to portray their determination in believing in such a matter despite the uncertainties of the pandemic situation.

The results of the analysis also indicate that scientific researchers do not prefer to assert their propositions with a high level of certainty where the topics discussed were fairly focused on either epidemiological study or therapeutics research. Notwithstanding the

study's small sample of research articles, there were still research foci that were addressed with absolute certainty by these experts. Firstly, the scale and severity of the pandemic undoubtedly had an influence on researchers' motivation to conduct studies. This is because studies on epidemiology were apparent to be acknowledged by researchers through all three degrees of certainty. Despite so, high certainty was discovered to be pertinent to matters that need some sort of enforcement, predominantly those that correspond with therapeutics. For instance, the following are the top three collocates that cooccur with the modal *need*.

	Word	Cooccurrences	Candidates	MI ↓
1	urgently	5	5	10.93
2	learn	4	11	9.48
3	leaders	3	27	7.76

Figure 8. Collocates of the modal *need*

In reference to Figure 8, the three collocates of *need*, namely *urgently*, *learn* and *leaders* manifest the semantics of conveying a message that is considered vital. At a time of uncertainty that requires strong determination in finding solutions, scientific researchers express their high certainty in propositions as a means to make emphasis on what needs to be considered by both the public and higher authorities. To illustrate, Figure 8 describes the sample concordance lines for the cooccurrence of *need* and *urgently*.

	Details	Left context	KWIC	Right context
1	doc#0	with replicating virus (both clinically evident and asymptomatic), are **urgently**	**needed**	to blunt the trajectory of the epidemic in the US.</s><s>Large differences be
2	doc#3	hese questions have not yet been definitively addressed, the world **urgently**	**needs**	a vaccine.</s><s>The pandemic is taking a heavy human toll,66 and large- s
3	doc#6	due to the unknown effects on mother and fetus, and these women **urgently**	**need**	an option for protection from COVID-19 when vaccine candidates are submit
4	doc#8	s in data collected. 5.</s><s>Robust clinical trials: Future effective therapies	**need**	to be **urgently** tested as part of clinical research trials.</s><s>Areas current

Figure 9. Sample concordance lines for *need* + *urgently*

From the concordance lines in Figure 9, the modal *need* is utilised by the researchers to stress on the significance of the subject matters being raised. The propositions are asserted with a high certainty of the truth of the propositions asserted. This might be a strategy employed by the researchers to mark their stance and make their readers pay particular attention to the proposition and thus, act accordingly. This is in line with Hyland's (2005b) argument of how a stance is a needed strategy to show one's work as being credible and reliable enough to gain acknowledgement by the addressees. Line 3 exemplifies the sense of cruciality of vaccines for pregnant women.

Excerpt 4

Figure 9, Line 3	Example in Context
need + urgently	"Pregnant women initially may not be eligible to receive COVID-19 vaccines due to the unknown effects on mother and fetus, and these women urgently need an option for protection from COVID-19 when vaccine candidates are not available to them."

The researchers give prominence to the issue of how the development of vaccine was lacking in terms of its side effects towards pregnant women and their unborn babies due to the scarcity of evidence. This matter is argued to be detrimental and the delay of providing a vaccine for them would cause death to both mother and fetus. The modal *need* well signifies their judgement, and readers can decipher the imperativeness of vaccines being provided as soon as possible. On account of this, it can be inferred that the portrayal of being certain by these researchers in relation to therapeutics research is regarded as purposeful to encourage the scientific community to expedite the process of finding effective drugs and vaccines.

Discussion

The current study examines scientific researchers' portrayal of their uncertainties in research articles that reflect various research foci. Their linguistic choices of epistemic devices were analysed, focusing particularly on modal auxiliaries. In the corpus analysis of the list of identified modals, the study observed that those that have the lowest epistemic modality value were used most frequently while those with the highest value of certainty were used the least. This confirms the state of uncertainty and the unwillingness of scientific researchers to fully commit to the truth of their propositions in relation to COVID-19 research. The research focus that was most often discussed by these researchers with the least amount of certainty was epidemiology where the analysis involves health and disease conditions, the findings of which contribute greatly to the shaping of science-based policies. Epidemiology was realised by all degrees of certainty but the lowest degree was the most pervasive. Even so, the certainty degree in which scientific researchers expressed their judgements about epidemiology was likely to be influenced by their intentions and the willingness for them to be held responsible for what they claim.

Based on the findings, it can be said that scientific researchers tend to use modals such as *may*, *could* and *might* to reference their lowest degree of certainty. Such portrayal of hesitancy in committing to their asserted propositions was mostly related to the topics of epidemiology where researchers intended to make guarded assumptions and flag other

possibilities contradicting what had been presented. With the virus that was rapidly evolving and spreading during the pandemic, these experts went through various challenges and difficulties due to the pressure from the public (Orso et al., 2020). In this study, the researchers' display of uncertainties was quite visible, particularly in the absence of relevant data. It departs from Orso et al.'s (2020) finding that scientists were more inclined to mask their uncertainties to appear more reliable via their truthfulness.

As regards portraying a higher degree of certainty with a probable epistemic modality value, scientific researchers demonstrated a tendency to conduct studies that involve clinical characterisations of the coronavirus. As in other similar studies (e.g., Dong et al., 2020; Tran et al., 2021), researchers were found to assert their propositions with more reliability through a display of a higher level of certainty. Nonetheless, absolute certainty was still avoided as the findings only suggested researchers' opinions and predictions of what would likely happen. Consequently, it would be unethical to make claims with high certainty with no sufficient support provided. Besides, this is a known factor of writing in the scientific community where claims should be made with proper hedging to minimise the prospect of imposition on the knowledge of others (Hyland, 1996). This fact is applicable even if the researcher is highly certain of his/her propositions (Hyland, 2005b), which possibly explain the lack of high certainty level epistemic devices found in the corpus data.

The researchers' depiction of high certainty was evident when they intended to make strong claims. This is regarded crucial during the pandemic as so much was at stake. It was found that the researchers were willing to dismiss the scientific writing convention and mark their certainty with high epistemic modality value when the propositional content involved matters that were urgent. In cases as such, the claims required special attention by the community and were related to epidemiology and therapeutics. These were foregrounded with strong modals to drive home the severity of the matters discussed, which include issues like the implementation of necessary health guidelines as well as the necessity of vaccines.

CONCLUSION

As a response to the alarming COVID-19 pandemic, it is evident that scientific researchers' linguistic choices in research articles were affected by the research foci. The types of epistemic devices used in relation to their epistemic modality values and degrees of certainty were not entirely as suggested in Halliday and Matthiessen's (2014) list of modal operators due to the absence of utilisation of some modals, particularly those that are negated. Further,

though epidemiology was widely discussed by scientific researchers via all three values of certainty, the manner in which their propositions were conveyed was influenced by the propositional contents. The frequent use of epistemic devices that bear the least degree of certainty indicates the unwillingness of researchers to make definite claims about the pandemic. This might be due to novel nature of the coronavirus at the time, and much of the research was still in progress. It was apparent that the uncertainties surrounding the virus still needed clarification, which led researchers to make assumptions of possibilities without affirmation of the truth. However, as the degree of certainty increased, researchers showed more determination with regard to committing to their propositions, though such occurrences were uncommon.

Finally, the small amount of data is obviously a limitation to the study. Thus, future studies should consider employing a larger corpus of COVID-19-related research articles while also taking into account epistemic devices from other lexicogrammatical items such as epistemic adverbs, adjectives, verbs as well as nouns. This will allow further exploration of the mechanism of each epistemic device in terms of its utilisation to express researchers' uncertainty in a face of pandemic.

REFERENCES

Adegbola, O. F. (2019). Points of view and modality in the discourses of homosexuality in selected Nigerian newspapers. *International Journal of Linguistics, Literature and Translation*, *2*(4), 80-88.

Akbas, E., & Hardman, J. (2018). Strengthening or weakening claims in academic knowledge construction: A comparative study of hedges and boosters in postgraduate academic writing. *Educational Sciences: Theory & Practice*, *18*(4).

Amsalem, D., Dixon, L. B., & Neria, Y. (2021). The coronavirus disease 2019 (COVID-19) outbreak and mental health: current risks and recommended actions. *JAMA Psychiatry*, *78*(1), 9-10.

Brown, T. S., & Walensky, R. P. (2020). Serosurveillance and the COVID-19 Epidemic in the US: Undetected, Uncertain, and Out of Control. *Jama*, *324*(8), 749-751.

Correia, M. I. T. (2020). Nutrition in times of Covid-19, how to trust the deluge of scientific information. *Current Opinion in Clinical Nutrition and Metabolic Care*, *23*.

Dong, M., Cao, X., Liang, M., Li, L., Liu, G., & Liang, H. (2020). Understand research hotspots surrounding COVID-19 and other coronavirus infections using topic modeling. *MedRxiv*.

Flowerdew, J., & Peacock, M. (Eds.). (2001). *Research Perspectives on English for Academic Purposes*. Cambridge University Press.

Halliday, M. A. K., & Matthiessen, C. M. (2014). *Halliday's Introduction to Functional Grammar*. Routledge.

Hengeveld, K., & Mackenzie, J. L. (2008). *Functional Discourse Grammar: A typologically-based theory of language structure*. Oxford University Press.

Hunston, S. (2002). *Corpora in applied linguistics*. Cambridge, UK: Cambridge University Press.

Hyland, K. (1996). Writing without conviction? Hedging in science research articles. *Applied Linguistics*, *17*(4), 433-454.

Hyland, K. (2005a). Metadiscourse: Exploring interaction in writing. Bloomsbury Publishing.

Hyland, K. (2005b). Stance and engagement: a model of interaction in academic discourse. *Discourse Studies, 7*(2), 173–192.

Iacobucci, G. (2020). Covid-19 makes the future of UK clinical research uncertain. *BMJ*, *369*.

Koffman, J., Gross, J., Etkind, S. N., & Selman, L. (2020). Uncertainty and COVID-19: how are we to respond?. *Journal of the Royal Society of Medicine*, *113*(6), 211-216.

Li, W., Sun, K., Zhu, Y., Song, J., Yang, J., Qian, L., & Wang, S. (2021). Analyzing the Research Evolution in Response to COVID-19. *ISPRS International Journal of Geo-Information*, *10*(4), 237.

Martín-Martín, P. (2008). The mitigation of scientific claims in research papers: A comparative study. *International Journal of English Studies*, *8*(2), 133-152.

Ngula, R. S. (2017). Epistemic modal verbs in research articles written by Ghanaian and international scholars: A corpus-based study of three disciplines. *Brno Studies in English 43*(2), 5-27.

Orso, D., Federici, N., Copetti, R., Vetrugno, L., & Bove, T. (2020). Infodemic and the spread of fake news in the COVID-19-era. *European Journal of Emergency Medicine*.

Rozumko, A. (2017). Adverbial markers of epistemic modality across disciplinary discourses: A contrastive study of research articles in six academic disciplines. *Studia Anglica Posnaniensia*, *52*(1), 73-101.

Simpson, P. (1993). *Language, Ideology and Point of View*. London : Routledge.

Teixeira da Silva, J. A. (2020). Misinformation in Covid-19 media and literature, with an emphasis on open data policies. *J. Advocacy Res. Educ.*, *7*(2).

Teixeira da Silva, J. A. (2021). COVID-19 research and publishing: Advance ambitiously, but cautiously, in 2021. *Indian Journal of Science and Technology*, *14*(10), 892-896.

Tran, B. X., Ha, G. H., Nguyen, L. H., Vu, G. T., Hoang, M. T., Le, H. T., ... & Ho, R. C. (2020). Studies of novel coronavirus disease 19 (COVID-19) pandemic: a global analysis of literature. *International Journal of Environmental Research and Public Health*, *17*(11), 4095.

Winiharti, M. (2012). The difference between modal verbs in deontic and epistemic modality. *Humaniora*, *3*(2), 532-539.

Yang, A., Zheng, S. Y., & Ge, G. C. (2015). Epistemic modality in English-medium medical research articles: A systemic functional perspective. *English for Specific Purposes*, *38*, 1-10.

Yeo-Teh, N. S. L., & Tang, B. L. (2021). An alarming retraction rate for scientific publications on Coronavirus Disease 2019 (COVID-19). *Accountability in Research*, *28*(1), 47-53.