



USE OF MODAL VERBS IN ACADEMIC WRITING: A COMPUTATIONAL FREQUENCY-BASED STUDY

Fakiha Arain

*MPhil Scholar, Department of English, National University of Modern Languages,
Faisalabad Campus. Email: fakihaarain23@gmail.com*

Dr. Aftab Akram

*Lecturer, Department of English, National University of Modern Languages,
Faisalabad Campus. Email: aakram@numl.edu.pk*

Abstract

This study examines the use of modal verbs in academic writing across two disciplinary domains: Natural Sciences and Social Sciences. The research aims to explore how modal verb choices contribute to the expression of stance, certainty, and interpersonal meaning within the framework of M.A.K. Halliday's Systemic Functional Linguistics. The study adopts a mixed-methods research design, combining quantitative corpus analysis with qualitative functional interpretation. A purposive sample of research articles from the two disciplines was compiled to construct two corpora representing Natural Sciences and Social Sciences texts. Corpus analysis was conducted using AntConc 4.3.1, which allowed the identification and frequency analysis of modal verbs through part-of-speech tagging using the CLAWS C7 tagset, focusing on central modal verbs, and they were subsequently analyzed in terms of their epistemic and deontic functions. The quantitative analysis provides an overview of the distribution and frequency of modal verbs in both corpora, while the qualitative analysis examines how these forms function within their disciplinary contexts. The findings indicate notable similarities and differences in modal usage between the two fields, reflecting distinct patterns of certainty, interpretation, and authorial stance. The study demonstrates that modal verbs operate as important interpersonal resources in academic discourse, shaping how writers present claims and negotiate relationships with readers across disciplinary boundaries.

Keywords: *Modal verbs, Academic writing, Modality, SFL.*

Introduction

Academic writing is a formalized form of communication used to create and share information within academic communities. It is typically defined as accuracy, caution, logically structured discussions, and arguments supported by empirical data. However, academic writing is never fully objective or impersonal; it often employs linguistic tools that convey certainty, possibility, appraisal, and obligation, thereby placing the writer in a position regarding his statements as well as the readers (Hyland, 2005). One of the most important of these resources is the modality. The concept of modality refers to the way in which a speaker or writer indicates his/her attitude to a suggestion using linguistic tools. It enables the language-users to communicate that something is possible, probable, necessary, or obligatory (Palmer, 2001). By allowing authors to tune the tenor of their argumentation rather than making it sound like a piece of unquestioned truth, modality is particularly important in academic discourse, where knowledge is habitually represented as provisional and liable to modification (Hyland, 1998).

In the linguistics discipline, modal auxiliary verbs include *may*, *might*, *must*, *should*, *can*, and *could*, which are the most commonly used modalities. These verbs depict a continuum of



commitment and change the nature of the meaning of the main verb. According to Huddleston and Pullum (2002), modal auxiliaries alter the truth value or force of a proposition by expressing the concepts associated with necessity and possibility. In this regard, modal verbs are used to encode different levels of epistemic and deontic commitments. Palmer (2001) proposes that there are two main modality, namely, epistemic and deontic. Epistemic modality deals with the likelihood or certainty of a proposition being true; it is the appraisal of the author of the information, inference, or probability. As an illustration, a statement like “The results can show a trend” will contain an element of uncertainty as opposed to solid certainty. On the contrary, deontic modality is concerned with obligation, acceptance, or need. Instead of dealing with different levels of truth, it reflects conventions, rules, or expectations. A statement like, “researchers should adhere to ethical standards” means that there is a duty that is based on disciplinary or institutional standards. In this way, deontic modality is concerned with the concepts of necessity and authority, and epistemic modality is concerned with knowledge and belief.

In academic writing, modality plays a variety of key roles other than simple categorization. It allows writers to indicate the levels of confidence, to weaken claims, to provide recommendations, to hedge claims, and to negotiate relationships with readers (Hyland, 2005). As a result, the use of modal verbs may also alter the formation of authorial position, as well as the determination of argumentative material. In the paradigm developed by M. A. K. Halliday, modality can be considered to be an element of the interpersonal metafunction of language (Halliday and Matthiessen, 2014). In this sense, language has ideational, interpersonal, and textual functions simultaneously. Since it conveys attitudes, judgments, and interactions between the writer and the reader, modality falls within the interpersonal aspect. Thus, modal verbs are not only the instrumental tools to build the meaning in the discourse, but also the grammatical structures.

In this research, the computational frequency-based approach will be used to conduct a critical analysis of modal verb use in scholarly documents. The outcomes of epistemic and deontic modalities in terms of frequency and distribution will be identified by reviewing academic articles of both Natural Sciences and Social Sciences. The study aims to explore how modal verb choices reflect differences in stance, certainty, and interpersonal meaning, providing insights into authorial positioning in academic prose within the framework of Systemic Functional Linguistics.

Research Questions

1. What is the frequency distribution of modal verbs in academic research articles from Natural Sciences and Social Sciences?
2. How do Natural Sciences and Social Sciences differ in their use of epistemic and deontic modal verbs?
3. How do modal verb choices in the two disciplines reflect differences in stance, certainty, and interpersonal meaning within the framework of Systemic Functional Linguistics?

Research Objectives

1. To determine the frequency distribution of modal verbs in academic research articles from Natural Sciences and the Social Sciences.



2. To examine how Natural Sciences and Social Sciences differ in their use of epistemic and deontic modal verbs.
3. To analyze how modal verb choices in the two disciplines reflect differences in stance, certainty, and interpersonal meaning within the framework of Systemic Functional Linguistics.

Significance of the Study

The study of modal verbs in academic texts is important because it clarifies how the authors express their attitude, beliefs, and interpersonal subtleties in their terms of speech. Knowledge of modal verb use is the way writers negotiate control over power, how careful or bold they should be in their speech, and how they locate themselves in relation to their statements and the audience. This research can help improve academic writing education for non-native speakers and inform corpus-based studies of language use in academia by highlighting disciplinary disparities in communication techniques across the Natural Sciences and Social Sciences.

Statement of the Problem

Despite the recognized importance of modality in academic writing, there is limited empirical research on the distribution and functional use of modal verbs across disciplines, especially in non-native academic contexts. Current studies often focus on native English writing or isolated disciplinary texts without systematically comparing epistemic and deontic usage. This gap makes it difficult to understand how authors in different fields convey certainty, obligation, and stance, and how these choices affect the interpersonal dimension of academic discourse. This study seeks to address this gap through a computational frequency-based analysis of modal verbs in Natural Sciences and Social Sciences research articles.

Literature Review

The ability of a single modal form to encode a variety of meanings based on the circumstances of the context makes modal verbs generally considered semantically complex. The taxonomy by Biber et al. (1999) roughly divides modality into three functional areas: possibility/ability (e.g., *can*, *could*, *may*, *might*), obligation/necessity (e.g., *must*, *should*), and prediction/volition (e.g., *will*, *would*, *must*). Khojasteh (2011) notes that such a variety of meanings habitually causes interpretation challenges, especially among English as a Foreign Language students who find it challenging to use the right modal forms in their academic work. The identical modal may indicate a variety of interpersonal connotations, and hence the status of modality as a linguistic tool that transcends grammatical action and is a pragmatic tool defined by a discourse situation. In 2018, the comparative study of hedging strategies was carried out by Demir (2016) in 200 English-language Teaching (ELT) research articles by Turkish and native English writers. The results showed that there was a statistically significant difference in the distribution of epistemic devices, especially modal verbs. Turkish writers mainly used modal auxiliaries, especially *can*, *may*, and *might*, but native writers used such alternative verbs and other indirect forms as *would*. The high rate of substitution of *may* by *can* by authors who do not speak English as a native language is possibly an indication of disciplinary or practical transfer phenomena. In his study, Demir highlights the importance of disciplinary positioning and rhetorical awareness on modal verb selection, which is significant to decipher the positions of authors of academic texts. Yang (2018) explored the modal verb occurrence in a professional corpus that consisted of published academic authors and a learner corpus that consisted of Chinese university students. Results



showed that modal verbs were more likely to be used by students, namely, the use of more frequent modal verbs, i.e., the usage of *can*, *will*, *could*, *would*, and a relative underuse of *may*. This gap poses the possibility that students can incorporate conversational constructs in their academic prose. The findings provided by Yang support the previous findings that have shown that novice writers are usually unable to tell the difference between the academic and colloquial styles.

Btoosh (2019) compared modal verb use by Arab university students with that of other non-native and native English speakers. Findings indicated that the Arab respondents made less use of epistemic terms like *may* and *might*, but attempted to use the modal verbs more often, like *must*, *can*, and *should*. The investigation explains this trend by the fact that some cultural values are inclined towards certainty-oriented expression. This result can be connected to the larger view that modality is a reflection of the subordinate socio-cultural norms that are being applied in discourse communities. In this way, the idea that modal verbs are used as markers of authorial identity and grounding of rhetoric is empirically supported in the study. Torabiardakani et al. (2015) relied on corpus-analysis software to investigate how modal verbs are used semantically in Iranian EFL compositions, thus going beyond works that merely counted frequency. The authors indicated that learners used possibility-oriented meaning less frequently, and preference was placed on ability, *can*, *could*, and obligation, *must*, and *should* meanings. This unequal distribution is a sign of a poor understanding of modal complexity in academic terms. Particularly, their approach based on the analysis of functions is especially acute, as they lay stress on semantic orientation rather than simply on the number of tokens.

In medical research articles, Dogan and Akbas (2021) examined epistemic positioning in particular subsections, results, and discussion. They found that modal boosters were relied on insignificantly, and that the use of modal auxiliaries to show the levels of certainty was extensive. This highlights the discussion section as a particularly fruitful place of modal analysis, as such a methodological implication of which has a direct bearing on the current investigation. Singh (2019) compared the Indian and British editorials in terms of Systemic Functional Linguistics (SFL). Singh emphasized the usefulness of SFL as an analytical instrument for identifying ideology coded in the linguistic framework. The research hypothesis is that an explicit claim of affective responsibility by the author forms the parameter of evaluation of modal aspects in SFL. The results indicated that in the efforts to distribute the blame on governmental participants, Indian editorial authors favored high-value modulated clauses and subjectively implicit orientations. On the contrary, British authors usually put the expectations and aspirations through medium value modulating sentences with objective implicit orientations. Such contradictions shed light on the differences in the status of authors and ideological orientation. Ahmad et al. (2019) compared the modal verb usage in Pakistani English newspaper editorials. The authors discovered that prediction modals, particularly the modal *will* and modal *would*, were most common, which proved an intensive usage of predictive framing in editorial speech. The scholars argued that Pakistani editorialists use various modal strategies in order to present institutional and personal perspectives as well as event coverage. The widespread usage of the *will* and the *would* are used to guide the reader towards specific points of view, as well as protect the identity of the author. Continuing this line of questioning,



Khalid (2013) examined the issue of authorial bias in editorials written in Scots English and discovered that modal meanings can be used as a measure of competence, agreement, and probability. In order to maintain the presence of the author and to improve the reader-author contact, the use of modal operators that would help to insert partial opinions into the text was deliberately used. Ntsane (2015) also emphasized the interactive aspect of editorial composition and claims that modality is an essential instrument of controlling the relationships between an editor and a reader. The modal devices are used to shape the views and perceptions of the readers and thus place the readers in the place of stakeholders in the argumentative discourse. Sadia and Ghani (2019) focused on the modal use in Dawn and The News editorials, but focused on the Pakistani media. Their findings also showed that they extensively used predictive modals, including, but not limited to, in both outlets, i.e., the use of the predictive modals, i.e., *will* and *would*. Nevertheless, the use of functional deployment differed: Dawn used obligation modals mainly to suggest solutions to the unresolved problems, and The News used predictive ones in order to predict the upcoming events. These trends imply the divergence in editorial policies and institutional positions.

The systemic-functional approach of Halliday (1994) provides the interpersonal role of language that can help writers attract the reader, maintain relationships, and impact views. The role of modal verbs as an interpersonal tool has already been studied in the literature (He et al., 2013), but the literature is limited in its coverage of their application in particular fields of study. As an example, abstracts of the articles in the Journal of Aerospace Sciences have been studied in the context of the aviation industry to demonstrate how modal verbs are used by the authors to help them communicate with each other (Linglit Journal, 2021). Similar analysis approaches may be used in other areas- humanities and social sciences- or to the situation of writing academic prose in Pakistan, to support the purpose of modal verbs in helping the reader to guide their interpretation.

As Palmer (2007) differentiates between propositional and event modality and then sorts them into epistemic, evidential, deontic, and dynamic, he proposed a typological system of mood and modality. Palmer asserts that evidential modality focuses on the fact that a thesis is supported by evidence, whereas epistemic modality reflects the opinion of the speaker in regard to the truth of an argument. Dynamic modality deals with inner capability or desire, and deontic modality deals with duties or allowances by outer factors. This taxonomy is commonly used to question the encoding of stance and interpersonal meaning on modal verbs in academic writing studies based on a corpus. As an example, Qun (2010) used corpus approaches to study the issue of modality in newspaper editorial texts and educational articles. The authors hypothesized that the leading position of the modals of *should* and *could* in the editorial ranking was determined by the necessity to balance assertion and attentiveness towards readers; as a result, the scientific research publication retention of subjective titles of knowledge, such as *I think*, was favored. These results will be helpful in studying the modal use in academic prose in Pakistan to demonstrate that modal verbs in academic prose are dependent on genre and communicative intent. In addition to deontic, epistemic, and dynamic implications, modal verbs in academic writings have myriad uses.

Hyland (1994) states that modal verbs are the most commonly used modes of portraying epistemic mitigation in EAP and English-Specific Purposes (ESP) textbooks, making them



crucial in hedging. The early studies by Huddleston (1971) and Ewer (1979) defined the first and second semantic types of modal verbs in scientific texts, distinguishing between capacity, possibility, necessity, prediction, and obligation. Thompson (2001) built on this basis by looking at modal use in theses of PhDs; he found eleven types of functional categories: the expression of prospective or requisite actions, hedging, characterization of subject attributes, legitimization, and using metadiscourse to guide the reader. This practical framework is an effective one in studying modal employment within academia. Although extensive studies have been done on the use of modal verbs by native speakers or by general learners of English, there are few comparative studies of modal verbs between native speakers and non-native speakers in academic writing (Hyland and Milton, 1997). In turn, the study of modal verbs in the Pakistani academic articles may close this gap by providing information on the way non-native authors manage the position, hedging, and interpersonal communication in academic writing.

The presence of modal verbs is a vital characteristic of academic prose, as the literature that has survived bears witness to this fact, since it helps the author to state positions, control discourses with the audience, and mitigate claims. The empirical studies involving the research of a variety of linguistic and cultural backgrounds prove that there is a significant difference in the use of modality by native and non-native authors; therefore, the challenge is maintaining the balance between epistemic, deontic, and dynamic functions. Although modal verbs have been the subject of significant research on the issues of academic writing, cross-cultural variation, and disciplinary distinction, there is an unavoidable gap in our knowledge about how Pakistani writers use these expressions, especially when one takes into consideration corpus-based, frequency-oriented studies. Modal verbs are significant in negotiating certainty and obligation, authorial identity, and rhetorical positioning. Therefore, by methodically examining the usage of modal verbs in academic writings, this study seeks to fill this gap and advance theoretical knowledge as well as practical consequences for academic writing training.

Methodology

Research Design

This study adopts a comparative corpus-based research design to investigate the use of modal verbs in research articles from Natural Sciences (NS) and Social Sciences (SS). The study follows a mixed-method approach, combining quantitative corpus analysis with qualitative functional interpretation. The quantitative component involves frequency-based analysis of modal verbs, while the qualitative component draws on Systemic Functional Linguistics (SFL) to interpret epistemic and deontic meanings. To ensure systematic and reliable analysis, the study utilizes corpus linguistic tools. The selected texts are processed using the CLAWS C7 POS tagger for accurate identification and tagging of modal verbs. The tagged corpora are then analyzed using AntConc version 4.3.1, which facilitates frequency counts, concordance analysis, and comparative examination across disciplines. The integration of these tools strengthens the reliability, replicability, and methodological rigor of the study.

Sampling Strategies

This study employs a purposive sampling technique to ensure the selection of relevant and high-quality research articles. A total of ten peer-reviewed research articles published in 2025 are selected, with five articles from Natural Sciences and five from Social Sciences. Although the total word counts of the two sub-corpora are 21,807 for NS and 37,378 for SS because of their



differences in size, normalized frequencies per 1,000 words were calculated to ensure comparability across disciplines. Only articles published in journals recognized by the Higher Education Commission (HEC) of Pakistan are included to maintain academic credibility and standardization. To ensure comparability across disciplines, only full-length empirical research articles are selected. The analysis focuses on specific sections of each article, including the abstract, introduction, results, discussion, and conclusion, as these sections are more likely to contain evaluative language and authorial stance. The literature review and research methodology sections are excluded to maintain consistency and to avoid sections that may contain formulaic or procedural language less relevant to modal expression.

Discipline	Number of Articles	Published Year	Word Count
Natural Sciences	05	2025	21,807
Social Sciences	05	2025	37,378

Table. 3.1: Overview of Selected Academic Research Articles by Discipline
Theoretical Framework

The proposed research is based on the Systemic Functional Linguistics (SFL) as proposed by Halliday (1994) and further developed by Halliday and Matthiessen (2014). The language in SFL is conceived as a social semiotic system; the meaning is constructed by the decision made in a system of linguistic options. Instead of viewing language as an unchanging system of abstract rules, SFL views language as a source of meaning-making, which is influenced by social contexts. The nature of academic writing is communicative and socially situated, and thus, SFL offers a proper framework for analyzing how writers create their stance, position of authority, and interpersonal positioning through modality.

The SFL theory indicates that language has the simultaneous accomplishment of three metafunctions: ideational, interpersonal, and textual. The current study puts the main emphasis on the interpersonal metafunction that implies interaction, evaluation, and negotiation of relationships between the writer and reader. In this metafunction, the central place is occupied by modality, which allows writers to portray their judgment, levels of certainty, obligation, probability, and inclination. In SFL, modality is construed to be between positive and negative polarity and presents levels of a middle degree of commitment. It allows authors to go beyond categorical statements and indicate shades of truthfulness or compulsion. Halliday identifies two broad categories of modality: modalization and modulation. Modalization is connected with probability and usualness (e.g., may, might, probably, usually) and is usually related to propositions (statements of fact). Instead, modulation is connected with obligation and inclination (e.g., must, should, have to, willing to) and has a connection with proposals (commands and offers). Within an academic context, modalization is especially important since academics are frequently required to make statements with caution, with some degree of likelihood, or with some form of alternative interpretation. Less common is modulation, which is found in recommendations, implications, and pedagogical suggestions.

Additional categorization of modality in SFL is determined in terms of value (high, medium, low), orientation (subjective versus objective), and polarity. Modes of high value (e.g., must,



certainly) are signs of high commitment, whereas modes of low value (e.g., may, possibly) are signs of tentativeness. Such shades play a vital role in scholarly writing, in which building authority and care rely on the balance between authority and care. The subjective orientation clearly means that the writer (e.g., I believe) is used, unlike the objective orientation, in which the assessment is given as impersonal (e.g., it is likely that), which is a common characteristic of formal research writing. In SFL terms, modality is more than grammatical, and it serves rhetorically and interpersonally. It codifies position and displays norms of discipline. Academic discourse requires that authors identify themselves with the existing body of scholarship, yet with a measure of epistemic humility; as a consequence, modal expressions are distributed and made stronger to play a role in the process of knowledge claims construction and negotiation within a discourse community.

In this context, modality is not only considered as a language system, but it is also a quantifiable discourse phenomenon. The analysis of the modal verbs and corresponding constructions in the selected sections of research articles is aimed at determining the patterns of interpersonal meaning-making. The study will explain the construction of stance in the various disciplines by mapping modal expressions onto SFL categories, including modalization, modulation, value, and orientation.

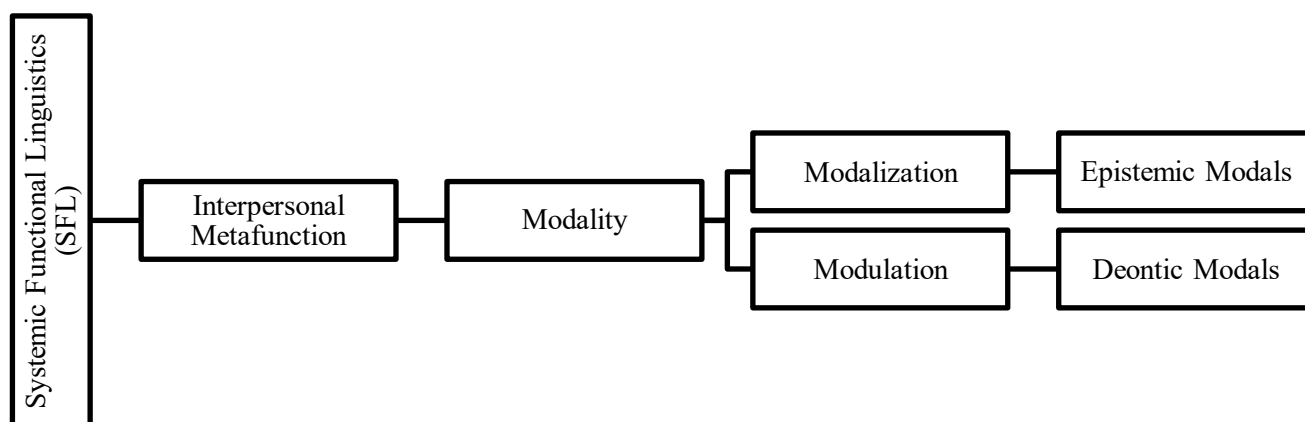


Figure 1: Theoretical Framework Based on Halliday's Interpersonal Metafunction
Data Analysis and Findings

This section of the research includes a quantitative study of the modal verb usage in the Natural Sciences (NS) and Social Sciences (SS) corpora. The analysis focuses on the general frequency of modal verbs in the two fields in order to determine the base tendencies of modal density and move on to functional and interpretive analysis under systemic functional linguistics (SFL).

Modal verbs were retrieved using the part-of-speech tags VM (central modal verbs) and VMK (modal catenatives) from the CLAWS C7 tagset, implemented through AntConc. The CLAWS C7 tagging system offers not only finer grammatical categorization, but it is often applied in corpus-based linguistic studies because it is quite reliable and descriptively precise. In this tagset,



VM is the core modal auxiliaries, including *can, could, may, might, must, shall, should, will, and would*, and VMK is core modal catenatives, multi-word modal constructions like *ought to and used to*. This tagset was used to ensure that modal forms in both sub-corpora were retrieved systematically and in a repeatable way. The examination revealed that central modal verbs (VM) were identified in the chosen research studies, but no modal catenatives (VMK) were found in the Natural Sciences or Social Sciences corpus. The lack of VMK forms suggests that modal catenatives are either not used or are extremely rare in modern-day academic research writing. The observation is in line with the stylistic rules of the academic discourse, in which more compact and syntactically thrifty modal auxiliaries are consistently preferred to multi-word modal structures, especially in formal written language.

A total of 52 tokens of central modal verbs in the Natural Sciences corpus, which is composed of 21,807 tokens, were found. On the other hand, the Social Sciences corpus, which includes 37,378 tokens, had 110 tokens of central modal verbs. In the corpus of the Social Sciences, modal verbs are more common at the level of raw frequency, but raw frequencies cannot be a legitimate source of cross-disciplinary comparison because of the size differentials in the size of the corpus. To deal with this problem, normalised frequencies per 1,000 words of both corpora were obtained. The findings show that modal verbs are used at a rate of 2.38 instances per 1,000 words in the Natural Sciences corpus and in 2.94 instances per 1,000 words in the Social Sciences corpus. These normalized values indicate that modal verbs are proportionately more common in the research articles in the Social Sciences than in the Natural Sciences articles, even when the corpus size is factored in. This quantitative distinction implies that authors in the Social Sciences use modal auxiliaries to a greater extent in building intellectual cases. The epistemological orientation of the two fields can be considered one of the possible reasons for this tendency.

The research of Natural Sciences is generally premised on empirical observation, experimentation, and measurement, which tend to promote restrained and carefully cast claims. Consequently, there is an opportunity to use modality more selectively. Conversely, the research of the Social Sciences often incorporates interpretation, assessment, and theoretical debate, situations where modal verbs are highly important in stating probability, requirement, and the author's position. Other studies in the past confirm this reading. To illustrate this, Hyland (2005) speculates that modal verbs are important indicators of position in scholarly writing, especially in the areas where argumentation and interpretation are the most prominent subjects. She observes that authors within the Social Sciences are more likely to use modal auxiliaries to bargain positions, recognize other parties, and locate themselves in relation to their audiences. The higher normalized frequency of modal verbs in the Social Sciences corpus in the present study aligns with this observation, suggesting a stronger interpersonal orientation and greater engagement with readers through modalized claims.

Discipline	Total Tokens	VM (Central Modals)	VMK (Catenatives)	Normalized Freq (per 1,000)
Natural Sciences	21,807	52	---	2.38
Social Sciences	37,378	110	---	2.94



Table. 4.1: Frequency of Modal Verbs in NS and SS Corpora
Epistemic and Deontic Distribution of Modal Verbs

The modal verbs found in both corpora were analyzed from the perspective of Systemic Functional Linguistics (SFL) to identify how modality operates in the interpersonal context of disciplinary discourse. The concept of modality in the SFL framework, as proposed by Halliday, can be understood as one of the interpersonal metafunctions of language, through which speakers and writers negotiate on the levels of certainty, probability, obligation, and inclination. Halliday differentiates between modalization, a concept that is related to probability and usualness (epistemic meaning), and modulation, which is associated with obligation and inclination (deontic meaning).

In scholarly writing, modalization usually takes the form of modal auxiliary verbs denoting probability, possibility, anticipation, or inferential reasoning, while modulation takes the form of modal auxiliary verbs denoting compulsion, advice, or need. To investigate the differences in modal meaning in terms of discipline, all modal verbs found in the Natural Sciences (NS) and Social Sciences (SS) corpora were divided according to the contextual concordance analysis. The modal verbs *may*, *might*, *could*, *would*, *will*, and *can* were defined as epistemic (modalization) because the lines of concordance in these modal verbs were mainly of the type possibility, inductive speech, prediction, or theoretical generalization. On the other hand, *must*, *should*, and *shall* became deontic (modulation), since they conveyed obligation, recommendation, procedure regulation, or normative position.

Quantitative report of Modalization and Modulation.

The statistics indicate that there is a strong dominance of the epistemic modality in both disciplinary corpora. Epistemic modal verbs in the Natural Sciences corpus explain 86.5 per cent, where deontic modal verbs explain only 13.5 per cent. The same distribution pattern is seen in the Social Sciences corpus, with epistemic modals taking 88.2 percent of the total modal usage and deontic modals taking 9.1 percent of the total. This distribution shows that both fields of academic research writing are dominated by modalization as opposed to modulation; that is, the modal is used mainly to express the levels of probability and epistemic position as opposed to obligatory and prescriptive authority.

Epistemic Modality Natural Sciences.

In the corpus of Natural Sciences, “*can*” (17 instances) and “*will*” (15 instances) are the most common epistemic modal verbs. Concordance analysis proves that “*can*” is very common in passive constructions, including *can be used*, *can be generalized*, *can be observed*, and *can be solved*. These constructions serve merely to articulate theoretical or rational possibility, but not permission. On the SFL scale, they modalize probability at an intermediate level of certitude. Putting the passive voice additionally diverts attention away from the author and to the processes or phenomena, and thus supports the impersonal feeling of scientific language. Modal “*will*” frequently occurs in anticipatory situations, especially when reporting the results of experiments or speculations about theories. Unlike in its conversational application, where “*will*” can be used to indicate intention, when writing science, it can often be used to indicate logical consequence or expected outcome, thus serving as an epistemic prediction and not a volitional one. The relatively low percentages of “*might*” and “*would*” indicate that authors of the Natural Sciences do use fewer highly tentative or hypothetic forms than authors of the Social Sciences. This trend



could be indicative of empirical bias in natural sciences research, where assertions are based on experimental evidence and therefore are made with controlled yet measured confidence.

Epistemic Modality of Social Sciences

The corpus of the Social Sciences shows an even greater dependency on epistemic modality, especially with high frequency of “*can*” (38 instances), “*could*” (18 instances), and “*would*” (14 instances). Of particular importance is the prevalence of “*could*” and “*would*” that so often indicate hypothetical reasoning, interpretative analysis, and counterfactual positioning. As an illustration, frequently presents interpretations of an analytical nature, which imply possible explanations but not necessarily conclusions. Similarly, *would* can be frequently used in argumentative situations to establish what is likely to happen provided that specific theoretical conditions are held. These forms enable authors to build arguments as plausible possibilities and not truths, thus bargaining interpersonal fit with readers. According to the SFL model, this use indicates a calculated use of modalization to control epistemic responsibility. Authors, by making claims look likely and not definite, leave room for possible alternatives and scholarly discussion. This especially applies to the field of research in Social Sciences, where a phenomenon is often a complex, context-specific, and subject to a variety of theoretical views.

Deontic Modality of Natural Sciences

The deontic modal verbs are used in the corpus of Natural Sciences quite rarely, which means that the manifestation of the obligation or a directive authority is not typical of this field. The most common modal is the “*should*,” and six instances of it were found in the research articles. Such cases are mostly related to the start of the discussion or conclusion parts, where the authors offer procedural recommendations or give directions on where the further investigation should head. The modality has to be used even less frequently: once, which indicates that categorical necessity or a high level of obligation is not so frequently expressed in empirical research articles. Similarly, there is little prescriptive or authoritative language use, as there is little usage of “*shall*.” In the context of the systemic functional linguistics (SFL) approach, such a low deontic modulation frequency indicates that writers in the Natural Sciences are not expected to establish themselves in the role of an authority figure and lay down rules or norms. Rather, the discussion focuses more on empirical rigour and methodological openness, and the procedural recommendations and suggestions are formulated in a way that does not affect the interpretative freedom of the reader. The few deontic formations are also consistent with the epistemological disposition of the Natural Sciences that emphasizes verification, repeatability, and well-defined experimental protocols at the expense of prescriptive or normative utterings.

Deontic Modality of Social Sciences

On the other hand, the Social Sciences corpus too shows an equally low proportion of deontic modals but has minor disciplinary differences. The modal “*should*” is used six times, which is similar to the situation in the Natural Sciences and is used mostly in the discussion and conclusion parts when authors suggest theoretical implications or the direction of future research. Social Sciences has slightly more instances of the use of the word *have to*, three, which may indicate that there are instances when there is a focus on necessity or obligation, more in terms of interpreting or doing. Just as in the Natural Sciences, there is almost no use of the term “*shall*,” which once again emphasizes the point that the use of authoritative and commanding language is not one of the key characteristics that define writing in the scholarly Social Sciences. However,



relative to the Natural Sciences, the Social Sciences corpus displays a marginally greater overall presence of hypothetical and epistemically nuanced forms, e.g., “could”, “would”, often interwoven with deontic expressions. Such a combination means that the authors of the social sciences often pair the recommendations or procedural suggestions with the reasoning that is rather probable or interpretative. In the light of SFL, this proves that modulation exists but is strictly controlled to negotiate using obligation in a way that acknowledges the interpretative space of the reader; thus, it portrays the interpersonal role that language plays in creating relational meaning, as opposed to providing instructions.

Cross-Disciplinary Comparison

Although both Natural Sciences and Social Sciences predominantly favor epistemic modality, several nuanced differences emerge when comparing the two disciplines. In the Natural Sciences corpus, authors exhibit a slight preference for deontic forms when providing methodological guidance, procedural instructions, or recommendations for experimental design. These deontic expressions are typically confined to discussion or conclusion sections, where the authors propose steps for future research or suggest optimal approaches to data collection and analysis. Such usage reflects the structured, precise, and objective nature of scientific writing, where recommendations are framed carefully to maintain a formal, evidence-driven tone rather than to assert authority over the reader. The scarcity of strong deontic forms, such as *have to* or *shall* reinforces the norm that Natural Sciences authors rely primarily on probabilistic and empirical reasoning rather than prescriptive statements.

In the Social Sciences corpus, the use of deontic modals is similarly infrequent but demonstrates distinctive characteristics. When employed, deontic verbs such as *should* and *have to* are often coupled with epistemic hedging forms like *could*, *might*, or *would*. This combination reflects the interpretative and analytical nature of Social Sciences research, where authors frequently negotiate theoretical stances, discuss alternative perspectives, or present cautious recommendations. Social Sciences authors rely on this interplay between deontic and epistemic modality to manage the uncertainty inherent in interpretative inquiry, signaling tentative advice or conditional guidance rather than categorical directives. The higher occurrence of hypothetical and inferential forms further illustrates the discipline’s orientation toward nuanced argumentation, reflective reasoning, and theoretical exploration.

Comparing the two disciplines highlights key contrasts in disciplinary epistemologies. Natural Sciences writing emphasizes procedural rigor, replicability, and methodological clarity, resulting in a relatively higher but still modest use of deontic forms for structured guidance. Social Sciences writing, in contrast, emphasizes negotiation, interpretation, and argumentative flexibility, resulting in deontic forms being intertwined with epistemic modal verbs to convey provisional guidance and measured suggestions. In both cases, the infrequent use of strong deontic forms underscores a shared academic convention: scholarly authors prioritize cautious, probabilistic reasoning over authoritative or prescriptive assertions, signaling the collaborative and interpretive nature of knowledge construction.



Category	Modal Verb	NS Frequency	SS Frequency
Epistemic (Modalization)	can	17	38
	will	15	14
	may	6	6
	could	4	18
	would	2	14
	might	1	7
Total Epistemic		45	97
Deontic (Modulation)	must	1	3
	should	6	6
	shall	0	1
Total Deontic		7	10
Grand Total (VM)		52	110

Table. 4.2: Epistemic and Deontic Distribution of Modal Verbs in NS and SS Corpora
Interpersonal Meaning in the Natural Sciences

In the Natural Sciences corpus, the most common modal verbs are “*can*” (17 occurrences) and “*will*” (15 occurrences), followed by “*may*” and “*should*” (6 occurrences each). Upon examination of concordance lines, it can be noted that these modals mainly encode logical possibility and predictive certainty, which is based on empirical procedure. As an example, in the extract about the production of chitosan, the text states that: “can be used in biomedical, agriculture, cosmetic, wastewater and food packaging applications,” and that “this waste could be used for the production of chitosan, a low-cost derivative with great economic value.” In such cases, both “*can*” and “*could*” are not used in a conversational sense of permission or speculation, but as codes of theoretical possibility using the chemical properties and experimental data. Within SFL, this represents modalization of probability at a median degree of certainty, where claims are presented as logically supported rather than subjectively asserted. In the same vein, the statement that “these techniques will be further optimized for pilot and industrial-scale production” demonstrates the predictive function of will in scientific discourse. In this case, the meaning of “*will*” is not personal intention but procedural development that is presumed based on the development of research. This kind of predictive modalization is an indication of a fairly high level of epistemic commitment and supports the forward-looking and goal-driven character of scientific reporting. The fact that highly tentative forms like “*might*” (1 occurrence) and “*would*” (2 occurrences) are rather rare indicates that the writers of Natural Sciences maintain controlled certainty. Their position is reserved but evidence-based, which is a measure between hedging and evidence. The high rate of passive forms, in combination with modal verbs, also depersonalizes statements, and the interpersonal emphasis is not on the author but on the processes that can be observed. From an SFL perspective, this means that modality in the Natural Sciences is used to create objectivity while recognizing the extent of probability.

Interpersonal Meaning to the Social Sciences

The Social Sciences corpus, in turn, presents a different modal profile with the most common forms of modal being the following: “*can*” (38 occurrences), “*could*” (18 occurrences), and



"would" (14 occurrences). Although the use of the word "*can*" is more evident in both fields, the role of the word in the SS corpus displays a more interpretative inclination. For instance, the text notes that "Advertisements can influence the worldview of people and are used as a tool to impart a certain type of ideology in society." Here, the word "*can*" refers to a general possibility, rather than an experimentally validated capability. It serves to present a claim of social impact, with the interpretation being flexible. Similarly, statements such as "Women can be seen carrying the water pipe on the shoulders" or "The content can be linked with the cultural values of society" frame observations as analytically constructed rather than empirically measured facts. The modal "*could*" serve as another supplement to this interpretative position. When the text suggests that shell waste "could be used for the production of chitosan" in NS, it indicates practical potential; however, in the SS corpus, could typically introduces theoretical or hypothetical positioning. Likewise, would frequently appear in projected or evaluative contexts, implying that certain portrayals "would challenge stereotypes" or "would influence viewers," thereby constructing claims as conditional and context-dependent. These modal choices within the context of SFL are modalization to a different degree of probability, which in many cases is at a lower level of certainty than in the Natural Sciences corpus. The extensive application of hypothetical forms indicates the argumentative and discursive character of Social Sciences research, in which interpretation, ideology, and social construction are felt as important analytical issues. Instead of making claims of stable results, authors negotiate meaning by use of tentative positioning and reader involvement.

Comparison Construction of Stance

Despite the overwhelming dominance of both fields in using epistemic modality (NS = 86.5%; SS = 88.2%), the nature of epistemic modal expression is an important index of disciplinary position. The Natural Sciences focus more on competence and forecasting "*can*," "*will*," and build knowledge as being empirically based and viable in its procedures. Quite on the contrary, the Social Sciences have shown more consistency of choice and evaluation forms "*could*," "*would*," indicating a dialogic and evaluative orientation. Deontic modal verbs are not common in either corpus, and "*should*" is mostly used as a form of recommendation. In Natural Sciences, such suggestions generally have to do with methodological rigor, and in the Social Sciences are frequently intertwined with epistemic hedging to weaken the prescriptive authority. This trend shows how both academic writers do not resort to imposition in their academic writings, but rather prefer most probable arguments. Within the SFL perspective, these results prove that modality is an interpersonal resource that is central to academic discourse. By using strategic choice of modal verbs, authors measure certainty, build disciplinary power, and control relational stance towards readers. The stance of the Natural Sciences is based on controlled prediction and logical ability, whereas the stance of the Social Sciences is based on the interpretive possibility and theoretical negotiation. Therefore, the modal verb selection is not a matter of grammatical preference only; the modal verbs represent different types of epistemological orientations and interpersonal relationships. With the help of viewing modality in terms of SFL, the readers can see that the stance, certainty, and engagement of the reader are formed systematically with the help of the disciplinary conventions.



Discussion

The findings of this research establish unique trends in modal verbs distribution and functional use in corpora of the Natural Sciences (NS) and Social Sciences (SS). In general, it can be seen that core modal verbs (VM) as defined by the C7 tagset of the CLAWS are being used systematically in both fields, but modal catenatives (VMK) like the ones with the prefix “*ought to*” and the suffix “*used to*” are simply nonexistent. This omission supports the claim that academic research writing prefers structurally compact and semantically precise modal forms and, thus, avoids less formal or narrative-oriented forms. The identified trend supports previous corpus-based studies of academic discourse, which have focused on the limited and traditionalized quality of modality in academic writing (Hyland 1998). Although the Social Sciences corpus (37,378 words) contains more modal verbs in raw numbers (110) than the Natural Sciences corpus (21,807 words; 52 instances), normalization shows that Social Sciences actually have a slightly higher modal density (2.94 per 1,000 words) compared to Natural Sciences (2.38 per 1,000 words). This suggests that writers in the social sciences tend to use modal expressions a bit more frequently. However, the difference is not very large, indicating that modality is an important linguistic tool in both fields.

When individual modal verbs are closely analyzed, issues are identified in terms of similarities and differences in the two corpora. Among the most frequent forms, one can find the words “*can*” (38 occurrences), “*could*” (18), “*will*” (14), and “*would*” (14) in the corpus of Social Sciences. These verbs are mainly epistemic, meaning that they are verbs of possibility, probability, or theoretical projection. Their high frequency is a manifestation of the interpretive and argumentative character of the Social Sciences research, when the authors often comment on the possible implications, theoretical options, and hypothetical situations. Conversely, in the Natural Sciences corpus, dominance over the use of “*will*” and “*can*” is made up of (15) and (17) occurrences; however, their application is mostly associated with procedural exposition, general scientific ability, or predictable results based on empirical evidence. Considering the Systemic Functional Linguistics (SFL) approach to the matter, these findings suggest that modality functions in the interpersonal metafunction rather than in the resources metafunction, where it is used as the means to negotiate certainty, stance, and control over the relationships between an author and a reader. The preponderance of epistemic modality in both fields suggests that academic authors would give more priority to depository knowledge building as compared to the claim of firm commitment. This is a trend that indicates a normative expectation of scholarly communication to shun over-generalisation and to make claims that are open to empirical testing or disproving.

This interpretation is further supported by the relative scarcity of strong deontic modals like the ones that have an “*must*” and “*shall*.” The corpus of the Natural Sciences includes only one instance of using the word *must*, indicating that the issue of imposing obligation is infrequent. Recommendations are instead constrained by the terms of moderation, like *should* or placed in terms of necessities of methodology rather than authority. Most of the Social Sciences corpus has a little more instance of “*must*,” but it is still restricted, and is often found within the larger frames of arguments, as opposed to being introduced as categorical imposition. The relative scarcity of the use of the form “*shall*” (only one time in SS; none in NS) supports the dropping usage of very formal or prescriptive modal forms in modern scholarly writing. These trends



indicate more epistemological disparity in the fields. In the Natural Sciences, the focus of research is normally on the area of replicability, predictability, and observability. As a result, modal verbs often signal logical consequence “*will*” or general capability “*can*,” reinforcing an impersonal and objective tone. Social Sciences research, on the other hand, deals more frequently with human behavior, social systems, and theoretical interpretation. Consequently, modal verbs such as “*could*,” “*might*,” and “*would*” play a greater role in constructing tentative interpretations and accommodating multiple perspectives.

Notably, the epistemic over deontic prevalence in both corpora speaks to a certain level of adherence to a requestable academic stance. Writers make use of modality not to encode the grammatical meaning only, but to define the level of commitment associated with propositions. By skillfully balancing between certainty and possibility, authors protect their credibility and, at the same time, encourage the participation of the readers. This tactful equilibrium of control with care is one of the main aspects of the development of the scholarly voice. As per SFL theory, modality is not just a grammatical category but rather a resource of substance in the performance of social meaning. The use of specific modal verbs by the authors has the effect of locating the authors relative to their claims, evidence, and disciplinary communities. The comparative results, thus, verify that, despite similarities in the academic conventions of both fields, slight differences in the epistemological orientation, as well as interpersonal approach, also exist.

Conclusion

The current research aimed at exploring the occurrence and the functional use of modal verbs in scholarly writing in two fields of study: the Natural Sciences and the Social Sciences. Three main questions were discussed in the research: the general frequency of modal verbs, the percentage of epistemic and deontic modalities, and modal options as indicators of differences of stance, certainty, and interpersonal meaning within the framework of the Systemic Functional Linguistics (SFL). The initial research question involved the occurrence of modal verbs in the chosen corpora of the two fields. The results indicate that modal verbs form a regular aspect of academic writing of both Natural Sciences and Social Sciences texts. Even though there is a greater raw number of modal verbs in the Social Sciences corpus (110 instances) than the Natural Sciences corpus (52 instances), when normalised, these two disciplines use modal verbs at similar frequencies. This implies that modality is a key aspect in organizing academic argumentation and asserting levels of certainty in academic disciplinary settings. The second research question was that to examined the allocation of the epistemic and deontic modality across the two corpora. It was found that the domination of epistemic modal verbs is enormous in both fields. In the corpus Natural Sciences, about 86.5% of the modal verbs are epistemic; only a minor percentage are deontic. This trend is more or less the same in the corpus of Social Sciences, where epistemic modality constitutes approximately 88.2 % of all the modal uses. Such is epistemic modality preference, which is indicative of the academic writing rules where the writers are more preoccupied with the articulation of probability, possibility, and careful consideration rather than the force of direct obligation or authority. The third research question was how modal verb choices indicate disciplinary variations in stance, certainty, and interpersonal meaning in the framework of SFL. The results show that despite the high levels of epistemic modality that are applicable in both fields, both of them make use of different modal verbs in their argument constructions. In the corpus of the Natural Sciences, modal verbs are



most common: can and will, which are often used to describe the possibility to perform something, the possibility of a process (procedural), and the results that would be predicted by empirical data. This trend is indicative of a certain degree of certainty based on observable data and test outcomes that are relatively stable. In comparison, all the modal verbs like “could” and “would”, which are generally related to hypothetical considerations, interpretative analysis, and theoretical projection, are more frequent in the Social Sciences corpus. These forms provide authors with the opportunity to make claims provisionally as well as to admit other possible interpretations, which is characteristic of a more discursive and argumentative research in the social sciences.

In the light of Systemic Functional Linguistics, these results prove that modality is mainly a task of the interpersonal metafunction and that it is an important linguistic tool to create the authorial stance and the regulation of the relations with the readers. When used with a prudent choice of modal verbs, academic writers can regulate the level of confidence they include in their statements and place their arguments within the requirements of their field. The fact that both corpora use epistemic modality as the predominant mode of expression shows that both are dedicated to constructing knowledge cautiously, and the disparities in modal use show the disciplinary commitment to linguistic strategies.

Recommendations for Future Research

The area of future research can be further elaborated with the help of the current study, as it can be analyzed using modal verb use in bigger and more varied academic-based corpora and in other fields, such as engineering and medical sciences, and humanities, in order to make a more comparative approach. Subsurface research might also explore modal verbs in various parts of research papers, including introduction, literature reviews, methods, and discussion sections, as modality tends to be used with different rhetorical functions in each section of a research paper. Furthermore, the consideration of cross-linguistic analysis, especially in the comparison of the English academic texts with the texts written in other languages, can provide some significant information on how cultural and linguistic elements affect the statement of stance and certainty. Such research would help in understanding modality as one of the interpersonal resources in academic discourse.

References

- Ahmad, M., Mahmood, M. A., Mahmood, M. I., & Siddique, A. R. (2019). Use of Modal Verbs as Stance Markers in Pakistani English Newspaper Editorials. *Online Journal of Communication and Media Technologies*, 9(1). <https://doi.org/10.29333/ojcm/5722>
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar Of Spoken And Written English*. Longman.
- Btoosh, M. A. (2019). Modals in Arab EFL learners' composition: A corpus-based approach. *Linguistics and Literature Studies*, 7(3), 100-109.
- Demir, C. (2018). Hedging and academic writing: an analysis of lexical hedges. *Journal of Language and Linguistics Studies*, 14(4), 74-92.
- Doğan, Z. N., & Akbaş, E. (2021). An exploratory study of epistemic stance in results and discussion sections of medical research articles. *Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 25(3), 1132-1150.



- Ewer, J. (1979). The modals in formal scientific discourse: function, meaning and use. Santiago, Chile, University of Chile, Department of English Research Report Mimeograph.
- Halliday, M. A. K., & Matthiessen, C. (2014). Halliday's introduction to functional grammar (4th ed.). Routledge.
- Halliday, M.K.A. (1994). An Introduction to Functional Grammar. London: Edward Arnold.
- He, Y., & Wang, H. (2013). A corpus-based study of epistemic modality markers in Chinese research articles. In Workshop on Chinese Lexical Semantics (pp. 199-208). Springer, Berlin, Heidelberg.
- Huddleston, R. (1971). The sentence in written English: a syntactic study based on an analysis of scientific texts. Cambridge: Cambridge University Press.
- Huddleston, R., & Pullum, G. K. (2002). *The Cambridge grammar of the English language*. Cambridge University Press.
- Hyland, K. (1994). Hedging in academic writing and EAP textbooks. *English for specific purposes*, 13/3, 239-256.
- Hyland, K. (1998). Hedging in scientific research articles. John Benjamins.
- Hyland, K. (2005). *Metadiscourse: Exploring interaction in writing*. Continuum.
- Hyland, K. and Milton, J. (1997). Qualification and certainty in L1 and L2 students' writing. *Journal of Second Language Writing*, 6/2, 183-205.
- Khalid, P. Z. B. M. (2013). Modality analysis of the newspaper articles about the Scottish Ship RMS Queen Elizabeth. *Academic Journal of Interdisciplinary Studies*, 2(9), 458- 461. <https://doi.org/10.5901/ajis.2013.v2n9p458>
- Khojasteh, L. (2011). *Modal auxiliary verbs used in Malaysian English language secondary school textbooks*. Thesis submitted to the School of Graduate University Putra Malaysia. PhD dissertation.
- Ntsane, M. F. S. (2015). The management of writer-reader interaction in newspaper editorials. *Ghana Journal of Linguistics*, 4(2), 108-123. <https://doi.org/10.4314/gjl.v4i2.5>
- Palmer, F. R. (2001). *Mood and modality* (2nd ed.). Cambridge University Press.
- Palmer, F. R. (2007). *Mood and Modality*. Beijing: World Book Publishing Company.
- Qun, Z. (2010). Modality and generic features in Chinese EFL writing. *Chinese Journal of Applied Linguistics*, 5(33), 40-51.
- Sadia, S., & Ghani, M. (2019). Modality in Editorials of Pakistani English Newspapers: A Corpus-Based Study. *International Journal of English Linguistics*, 9(1), 144. <https://doi.org/10.5539/ijel.v9n1p144>
- Singh, C. J. (2019). Modality and Ideology - A Systemic Functional Linguistics Study. *Language in India*, 19, 56-65. <http://www.languageinindia.com/june2019/index.html>
- Thompson, P. (2001). A pedagogically-motivated corpus-based examination of PhD theses: macrostructure, citation practices and uses of modal verbs. Reading.
- Torabiardakani, N., Khojasteh, L. & Shokrpour, N. (2015). Modal auxiliaries and their semantic functions used by advanced EFL learners. *Acta Didactica Napocensia*, 8(2), 51-60.
- Yang, X. (2018). A corpus study of modal verbs in Chinese learners' academic writing. *English Language Teaching*, 11(2), 122-130.