

Modality and Evidentiality in English: A Semantic-Pragmatic Analysis of Speaker Stance and Information

Muhammad Umair Ashiq

BS English Language and Linguistics, University of Management and Technology, Lahore, Pakistan

Corresponding author: Muhammad Umair Ashiq (Email: umairashiq098@gmail.com)

Received: 29/01/2026, Revised: 26/03/2026, Accepted: 05/06/2026

ABSTRACT

Modality and evidentiality are two similar semantic categories by which English speakers' express attitude and knowledge about the propositions they express. Modality conveys possibility, necessity, obligation, and permission in various degrees, and evidentiality indicates the origin and/sentiment of the information presented. This article reviews recent scholarship (2019-2026) to reevaluate the theoretical links between these two categories in English and looks at their functioning in authentic discourse (both academic texts, news reports, and political speeches). The study takes a qualitative approach to research, based on documents, and examines studies on modality and evidentiality that rely on a corpus, discourse analysis, and typological analysis, as well as those from the foundational frameworks that have been offered by earlier researchers like Palmer, Bybee, Aikhenvald and others. The review demonstrates that English does not have an obligatory grammatical evidential system, but instead its speakers and writers use a wide range of lexical and syntactic evidential devices, mostly modal auxiliaries, evidential adverbs and reporting verbs, at times merging both the two functions into one expression. The discussion also demonstrates that these patterns are sensitive to register, genre and communicative context: hedging is increased in times of collective uncertainty, evidential marking is systematic in spoken and written academic discourse and second language writers show over- and under-use patterns when compared to first language writers. The article finds that modality and evidentiality are analytically separable but functionally dependent systems that continue to be influenced by new contexts of discourse, such as news texts in the pandemic era, political communication, and texts produced by artificial intelligence, and that future research should strive for larger and more varied corpora and for cross-linguistic comparison for a more detailed and nuanced theoretical picture.

Keywords: Modality, Corpus Linguistics, Discourse Analysis, Evidential Markers, Hedging, Speaker Stance

1. Introduction

Each assertion implies something about the speaker's level of confidence and how he or she knows what he or she is saying. The two dimensions, certainty and source, are routinely signalled by an overlapping set of grammatical and lexical choices, which have traditionally been the focus of the study of modality and evidentiality in English. Modality, in a general way, refers to the linguistic expression of possibility, probability, permission, obligation, and necessity: to assert a proposition as possible, probable, permitted, or obligatory, or as required, rather than as a plain fact. The linguistic encoding of the source of information underlying an assertion, by contrast, is evidentiality, which distinguishes, for example, a claim based on direct perception from one based on hearsay or inference. These two categories are conceptually distinct but there is significant overlap in English: a single modal verb like *must* can express both a high degree of epistemic certainty and an (inferred) basis for that certainty.

Modality and evidentiality have become a focus of recent interest due to the fact that the fields of corpus linguistics, discourse analysis, and computational text analysis have all become interested in how speakers and writers deal with issues of certainty and responsibility in authentic communicative situations. The COVID-19 pandemic in particular led to a significant amount of public discourse in which writers had to constantly balance what was known, what was suspected, and what was not, creating new opportunities for empirical studies concerning epistemic and



evidential marking. Meanwhile, a new precision of quantifying the use of modals and evidentials in registers, disciplines and first language backgrounds has been made possible through advances in corpus methodology, which is unavailable with earlier, more theory-driven accounts of the two categories.

This article reflects on this recent wave of studies, contextualizes them within the classic theoretical frameworks of Palmer, Bybee and Aikhenvald, and aims to shed light on the current understanding of the mechanisms of modality and evidentiality in English discourse and the trajectory of the field.

1.1. Significance of the Study

Beyond being of theoretical interest to linguistics, understanding how modality and evidentiality function in the present-day English discourse is of practical importance. To design effective instruction, language teachers and materials writers must get a clear and current understanding of where and when these kinds of resources are used and how they are used across registers, so that they can teach L2 writers how to use them appropriately and calibrate their level of certainty and source attribution. A bit of knowledge about the role of evidential and epistemic markers in constructing a sense of reliability is useful for journalists, editors, and media literacy teachers alike, particularly when it comes to reporting in high-stakes areas like public health, where the line between fact and inference can impact how audiences act. To convey these meanings from one language to another, translators and cross-linguistic communication specialists should have a keen sense of the nature of the relationship between English's lexically diffuse system of modal and evidential marking and the more grammaticalized systems found in many other languages. In conclusion, in the context of the growing use of automated and AI-generated texts in everyday communication, computational linguists and natural language processing researchers require more sophisticated models of modality and evidentiality based on empirical observations of the use of the phenomena in real English texts. This article brings together the last few years' research on these questions and is intended to be helpful to all these groups.

1.2. Research Objectives

The two main goals of this study are:

- Analyse and summarise the literature from 2019 to 2026 on the semantic categories of modality and evidentiality in English, focusing on their manifestation in academic, journalistic and political discourse.
- To discuss, based on this recent literature, how modality and evidentiality function in practice, that is, the degree to which they can be jointly or independently expressed by different discourse contexts and populations of speakers.

1.3 Research Questions

2. What can study from 2019 to 2026 tell us about the linguistic forms, functions and discourse distribution of modality and evidentiality in the English language?
3. What are the effects of modality and evidentiality in specific contexts of discourse, for instance in news reporting about the pandemic, academic writing and political speech, and what factors (register, genre, first-language background) influence this interaction?

2. Literature Review

As far as possible, the literature surveyed in this section will be limited to works published since 2019, with brief references to the classical theoretical frameworks that will be compared to the newer works. The review is structured thematically according to four strands of recent research related to modality and evidentiality: theoretical research on the relationship between modality and evidentiality, corpus-based research on epistemic modality in academic writing, discourse-analytical research focused on evidentiality in news, political and diplomatic discourse, and typological and grammaticalization research on the development of evidentiality in English.

2.1 Theoretical Foundations

The classical studies of Palmer (2001), Bybee (1985) and Aikhenvald (2004) showed that modality and evidentiality are related but different categories of semantics, the former dealing with degree of certainty and the latter with source of information. Recent research has modified and clarified this relationship, but not challenged it. In Mélaç

(2022), the author traces the historical grammaticalisation of evidential meaning in English, demonstrating that English did not develop an obligatory evidential paradigm like Quechua or Tibetan, but that some constructions (e.g., the evidential use of *seem* and *appear*) have undergone a process of gradual grammaticalisation that brings them closer to the function of evidentials in languages with dedicated evidential morphology. Furthering this line of enquiry, Mélaç (2024) compiles cross-linguistic research into ways in which modal expressions can over a period of time evolve into evidential expressions and the other way around, and concludes that modal expressions and evidential expressions should be interpreted as historically and functionally interconnected rather than as complete independent subsystems.

2.2 Epistemic Modality in Academic Discourse

There has been a wealth of recent corpus-based studies of the expression of epistemic modality in academic writing, sometimes in a comparative study of first-language (L1) and second-language (L2) writers. Drawing on a comparison between first-language English writers and first-language Chinese writers of English academic prose, Xie and Patten (2025) found that writers with a first-language Chinese background used a less diverse range of epistemic lexical verbs and were not very sensitive to the nuances of certainty these verbs can express, which has implications for teaching academic writing to this group. In a similar vein, Bozdağ, Morris, and Mo (2024) used Bayesian probabilistic modelling to examine the use of English modal verbs in L2 writing, comparing Turkish and Chinese L1 learners, and found that although there were no major differences in the general modal preferences between the two L1 groups, the topic-specific effects resulted in significant differences in modal choice, implying that modal verbs in L2 writing are determined by both L1 and by the communicative demands of the specific writing task.

To complement these studies that focus on the L2 learner, Wang (2024) did a corpus comparison study of evidentiality in spoken and written academic discourse involving personal, shared, and reporting evidentials. The study revealed that writers preferred the use of the shared evidentials and reporting evidentials, while speakers preferred the use of personal evidentials, which is commensurate with the less immediate and more accountable nature of the written academic registers and more interactive and self-oriented nature of spoken academic registers. The studies as a whole suggest that epistemic and evidential marking in academic discourse is not a monolithic phenomenon, but rather shows a systematic variation based on the mode of communication (spoken or written), first-language background and disciplinary conventions.

2.3 News, Political, and Diplomatic Discourse

But since then, there has been research that has explored the use of evidentiality in public discourse, where the stakes of source attribution may run higher. A quantitative comparison of the epistemic and evidential expressions in 100 Guardian newspaper articles from April 2020 (when there was a lot of uncertainty during the pandemic) and in early 2022 (when there had been a lot of medical information) was done by Alonso-Almeida (2023). The study revealed that the former subcorpus had significantly more expressions of weak commitment (e.g. *might*, *perhaps*, *apparently*) and strong commitment (e.g. *know*, *clearly*, *surely*) – both were indicative of a greater discursive need to distinguish between the known from the unknown in a period of collective uncertainty. This finding exemplifies the systematic changes in the density and distribution of modal and evidential marking that occur in public facing discourse as a result of outside events.

For political and diplomatic communication, Xu (2022) introduced a new framework of evidentiality into three categories: personal, shared, and other-source, and applied it to the data of political speeches of the last three American presidents. This analysis revealed that the speakers deliberately choose which of these evidential categories to use to either control or minimize their verbal responsibility and to legitimate or normalize specific ideologies; specifically, the use of shared-source evidentials is found to be a means of calling upon a presumed authority or collective knowledge. The study shows that evidential choice in high-stakes public discourse is often not a neutral function of the information source, but rather a rhetorical and ideological one.

2.4 Classification Frameworks and Modal Verb Corpora

In recent research another focus has been to improve the classificatory apparatus for the description of evidential and modal expressions beyond the traditional three-way split: sensory / hearsay / inferential. In the subsequent research that aims to go beyond the classical typological classification to a more discourse-oriented one, Xu's (2022) threefold system of personal, shared and other-source evidentiality has attracted attention, because political and diplomatic discourse often draws on knowledge that is collective or institutional and is not necessarily

aligned with individual sensory or inferential experience. Meanwhile, corpus-based research on modal verbs including the Bayesian modelling approach by Bozdağ, Morris and Mo (2024) has shifted away from mere frequency count to probabilistic and multivariate methods which can isolate the independent contribution of different variables such as first language background, topic and register in the choice of modals. This methodological turnaround is part of a wider trend in the literature for the 2019–2026 period in which the quantitative sophistication of corpus analysis is increasingly possible due to the increased availability of large, well-annotated academic, journalistic and learner corpora of English.

Over the period 2019–2026, the literature shares in common three key points that are defined by classical theory, yet also provides a significant amount of empirical and contextual detail. First, modality and evidentiality in English are not yet lexically or syntactically fixed and mandatory, but some new developments in grammaticalization research indicate that some of the evidential markers may be in the process of becoming more systemically used. Second, the use of markers of epistemicity and evidence is very dependent on the register, mode (oral vs. written), first-language background, and disciplinary conventions, as revealed by corpus studies of academic discourse. Third, evidential and epistemic options are often strategic decisions in public and political discourse, which may be connected to the communicative constraints of a certain historical situation (e.g., a public health crisis), but also to the ideologies and intentions of the discourse producer. Together, these findings provide impetus for the methodological approach and discussion presented in the remainder of this article.

3. Methodology

3.1. Research Design

This study is conducted with a qualitative research design which is based on documents and is appropriate for synthesis-oriented research on a semantic-pragmatic phenomenon. The study is not based on new corpus or experimental data but is a structured and thematic review and synthesis of the peer-reviewed literature on modality and evidentiality in English between 2019 and 2026, against the background of classical theoretical frameworks that are still dominant in the field.

3.2. Data Collection

Data collection was through a search strategy. The sources were identified based on the key words modality, evidentiality, evidential markers, hedging, deontic modality and epistemic modality, combined in various ways, and searched using databases and search engines focused on academic publications, including ScienceDirect, PubMed/PMC, ResearchGate and Frontiers; as far as possible, only publications from the period 2019 to 2026 were used. Peer-reviewed journal articles and book chapters with English language data received priority; other studies that used a corpus-based or discourse-analytic methodology were given special consideration as these methods are currently in vogue for empirical research on the topic.

3.3 Inclusion Criteria and Thematic Coding

Inclusion criteria and Thematic coding were used. Studies were selected that (a) included modality or evidentiality as a focus of analysis or both; (b) analyzed English language data, either exclusively or comparatively; and (c) were published between 2019 and 2026 (a few early studies were included if they were cited by the recent literature on a theoretical basis). The studies that met these criteria were then coded thematically based on the discourse domain (academic writing, news discourse, political/diplomatic discourse), the linguistic focus (modal verbs, evidential adverbs, reporting structures) and the methodological approach (corpus-based quantitative analysis, qualitative discourse analysis, typological/historical analysis). This thematic coding was used to organize the literature review and served as the categories for analysis in the discussion which follows.

3.4 Sampling

The following information was extracted and recorded for each of the study that met inclusion criteria: publication year, domain of discourse investigated in the study, the source of the data (e.g. newspaper corpus, learner corpus, political speech corpus), the number of samples (if reported), modal or evidential class examined, and the major findings relevant to the study questions guiding this paper. These details were then clustered in the four thematic strands presented in the beginning of the literature review—theoretical foundations, academic discourse, public and

political discourse, and typological/grammaticalization research so that patterns within and between clusters were systematically identified. This procedure was not as formal as a systematic-review but gave a consistent and transparent basis for the synthesis and discussion that follows.

3.5 Analytical Framework

The synthesis is based on a hybrid analytic framework that merges the classical contrast of epistemic modality and evidentiality (Palmer, 2001; Aikhenvald, 2004) with the more discourse-focused typologies presented in recent work, including Xu's (2022) typology of personal, shared, and other source and Wang's (2024) typology of personal, shared, and reporting for spoken and written academic discourse. This hybrid approach enables the categorization of the findings in the studies to be coded in terms of source type (personal, shared/institutional, reported, inferential), register (academic, journalistic, political/diplomatic) and communicative mode (spoken versus written), thus allowing the cross-study synthesis that is presented in the discussion section.

3.6 Limitations

This is a document-based synthesis and as such relies on the range and quality of studies that were located using the search strategy outlined above and does not incorporate a full systematic-review protocol with pre-registered search terms or inter-rater reliability checks for study selection. The review also privileges English language, corpus-based scholarship found in the databases searched and may thus underrepresent relevant scholarship in other languages and/or published in journals that are not well indexed by the search engines employed. Such restrictions are logical in a study of this kind, which is both exploratory and synthesis oriented, and are discussed below in the discussion section.

4. Discussion and Results

4.1 The 'Persistent Overlap of Modality and Evidentiality'

One of the common findings from the literature reviewed is that modality and evidentiality in English remain to be very much overlapping systems and are not neatly distinguishable systems. M  lac's (2022, 2024) grammaticalization studies demonstrate that this overlap has a historical dimension: evidential English meanings have often grown out of, and parallel to, modal and aspectual English meanings, and the distinction between the two types of constructions is not of a fixed nature. This diachronic view can account for the fact that synchronic studies of corpora (see the studies mentioned above) so often reveal a single lexical item (modal verb, adverb or verb of appearance) to have both epistemic and evidential meaning. The bottom line is that any analytical work that considers modality and evidentiality to be completely different variables may be mistaken in representing how these meanings are packaged in English use.

4.2 Mode of Communication

Wang's (2024) study comparing written and spoken academic discourse showed that the mode of communication plays a strong role in determining the choice of evidentials: shared or reporting evidentials were more common in written academic discourse, which tend to shift the burden of providing the evidence away from the individual writer, while personal evidentials were more common in spoken academic discourse, which foreground the speaker's own stance. This is consistent with well-established assumptions of the disparities between spoken and written discourse in terms of accountability, and the recent evidence in the corpus provides new empirical precision. In the same vein, the comparison of the April 2020 and 2022 subcorpora in the comparison of Alonso-Almeida (2023) reveals that register differences are influenced by contextual and historical factors: The same register (newspaper discourse) can exhibit significantly different densities of epistemic and evidential marking, depending on how much collective uncertainty is present at the time of writing. These findings taken together indicate that register and communicative context should not be seen as background variables, but as central explanatory factors in any explanation of the use of modals and evidentials.

4.3 First-Language Background and Pedagogical Implications

The studies reviewed above, all of which focus on the learner, confirm the following finding: writers of English as L2 and L1 make systematic differences in their use of epistemic devices as indicated by the different types of studies. Xie and Patten (2025) observed that L1 Chinese writers had a lower number of fine-grained degrees of certainty in their epistemic lexical verbs than L1 English writers, indicating a lack in their productive knowledge of such distinctions. Bozdağ, Morris and Mo (2024) revealed that generally, there was no significant difference between the two groups of learners (Turkish and Chinese) in terms of their use of modal verbs; however, these differences were notably significant when the writing topic was kept constant, suggesting that first language transfer effects do not work in the same way. All these findings add up to the pedagogical implication shared by much of the literature that the teaching of modality and evidentiality should leave behind the tradition of focusing on modality and individual modal verbs in isolation and instead treat the system of epistemic/evidential resources, which includes not only the individual modal verbs, but also lexical verbs, adverbs and reporting structures, as an interconnected set of stance-marking devices.

4.4 Strategic and Ideological Dimensions in Public Discourse

Xu (2022) has conducted a study on evidentiality in political discourse and finds that in high-stakes public discourse, the selection of any of the evidentiality types, be it personal, shared or other-source, is unlikely to be random. Speakers seem to have a conscious choice in the use of evidential categories to limit their own liability and to promote certain types of ideologization, especially when using shared-source evidentials, which evoke a sense of common or assumed knowledge, and that may be used for legitimizing purposes. This finding builds on how evidentiality has been discussed in the previous literature, which was mainly descriptive, and proposes that future studies of modality and evidentiality in public discourse should be sensitive to both the form and the political and rhetorical uses of evidentiality markers.

4.5 A human-AI collaborative text: Emerging Directions

While not yet the focus of systematic study in the 2019–2026 period under review, several of the corpus-based studies cited above mention in passing how important large language models are for the research on modality and evidentiality, as these models produce vast amounts of English text that may have systematic differences from human writing with regard to how modality and evidentiality are marked. An open question asked by this literature is whether existing taxonomies of evidentiality and epistemic modality that were developed mainly to describe human speech and writing, are applicable to automatically generated texts, or whether new taxonomies are required to describe the way that automatically generated systems express or fail to express information source and confidence. This is an exciting field for future research that can be built upon the research foundations presented in this article.

When read together, the results surveyed in this section confirm the view that modality and evidentiality are one and the same, that is, one complex, multifaceted system of resources to mark stances whose use is dependent on at least four factors: the historical grammaticalization pathways available in the language, the register and communicative mode of the event, the first-language background and proficiency of the speaker/writer, and the strategic or ideological requirements of the particular discourse context. It is overall reasonably consistent with the classical distinction between epistemic modality and evidentiality set up by Palmer, Bybee and Aikhenvald, but much more empirically detailed, and it suggests a research agenda that will use a combination of corpus-based, discourse-analytic and typological methods to capture the full complexity of how the English speaker signals what they know, and how confident they are in knowing it.

5 Conclusion

This article has surveyed studies of modality and evidentiality in English since 2019, and positioned recent, corpus-based and discourse-analytical research in the light of the classic theoretical distinction between epistemic commitment and marking an information source. The review reveals that this field of study is far from being exhausted, but persists as a response to new discourse contexts—most prominently the time of COVID-19, which has yielded an unusually large amount of data on the response of epistemic and evidential marking to times of collective uncertainty—and as a response to new methodological tools, especially large-scale corpus analysis and probabilistic modeling. Modality and evidentiality were shown to be conceptually distinct yet functionally overlapping systems, the combination of which is determined by register, mode of communication, first language background and the

strategic needs of public and political discourse. The results have implications for language pedagogy: the teaching of epistemic and evidential resources should not be approached from a purely grammatical perspective, but as part of an integrated system; and for discourse analysis and translation, where sensitivity to the writer's/speaker's epistemic and evidential stance is crucial to accurate interpretation. For future research, it would be desirable to have larger and more linguistically diverse corpora, systematic cross-linguistic comparison with languages with grammaticalized evidential systems, and more careful consideration of how new forms of communication (such as machine-generated and human-AI collaborative text) are affecting the ways in which certainty and information source are expressed in English.

It is important to recognize that the image of the situation here is incomplete. The studies reviewed were selected as a targeted rather than exhaustive search of major databases and search engines and the review admits English language corpus-based research that is well indexed in these resources; more specialized research published in other languages or in regionally based journals may not be well represented. Furthermore, due to the fact that the review is based on findings from studies that have used different corpora and different procedures for analyzing modality and evidentiality, the comparisons presented here are only indicative of trends in the field and not a statistical, integrated meta-analysis. Despite these constraints the overall picture that emerges; that a system which is lexically diffuse but discourse sensitive and whose use is determined by both historical grammaticalisation and register factors, but also by first-language background and communicative strategy, is coherent and based on empirical evidence, and thus constitutes a useful update to the classical theoretical account of modality and evidentiality in English.

Declarations

Conflict of interest	The author declares no conflicts of interest.
Funding	This research received no external funding.
Data availability	Data are available from the corresponding author upon reasonable request.
Use of AI tools	No AI-assisted tools were used in the preparation of this manuscript.
Author contributions	Conceptualization, M.A.; methodology, M.A.; software, M.A.; validation, M.A.; writing—original draft preparation, M.A.; writing—review and editing, M.A.; supervision, M.A.; project administration, M.A.

References

- Aikhenvald, A. Y. (2004). *Evidentiality*. Oxford University Press.
- Alonso-Almeida, F. (2023). Knowledge and belief in the times of COVID-19: A comparative analysis of epistemicity in English newspaper discourse of two stages of the pandemic. *Applied Corpus Linguistics*, 3(2), Article 100054. <https://doi.org/10.1016/j.acorp.2023.100054>
- Bozdağ, F. Ü., Morris, G., & Mo, J. (2024). A Bayesian probabilistic analysis of the use of English modal verbs in L2 writing: Focusing on L1 influence and topic effects. *Heliyon*, 10(6), Article e28701. <https://doi.org/10.1016/j.heliyon.2024.e28701>
- Bybee, J. (1985). *Morphology: A study of the relation between meaning and form* (Typological Studies in Language No. 9). John Benjamins. <https://doi.org/10.1075/tsl.9>
- Chafe, W. (1986). Evidentiality in English conversation and academic writing. In W. Chafe & J. Nichols (Eds.), *Evidentiality: The linguistic coding of epistemology* (pp. 261–272).
- Chung, S., & Timberlake, A. (1985). Tense, aspect, and mood. In T. Shopen (Ed.), *Language typology and syntactic description: Vol. 3. Grammatical categories and the lexicon* (pp. 202–258). Cambridge University Press.
- Mélac, E. (2022). The grammaticalization of evidentiality in English. *English Language & Linguistics*, 26(2), 331–359. <https://doi.org/10.1017/S1360674321000101>

- Mélac, E. (2024). The links between evidentiality, modality, and grammaticalization. *Studies in Language*, 48(3), 513–542. <https://doi.org/10.1075/sl.24011.mel>
- Palmer, F. R. (2001). *Mood and modality* (2nd ed.). Cambridge University Press. <https://doi.org/10.1017/CBO9781139167178>
- Portner, P. (2005). *What is meaning? Fundamentals of formal semantics*. Blackwell.
- Saeed, J. I. (2003). *Semantics* (2nd ed.). Blackwell.
- Wang, C., & Li, J. (2024). Comparative analysis of evidentiality in spoken and written academic discourse. *English Language Teaching and Linguistics Studies*, 6(3), 62–78. <https://doi.org/10.22158/eltls.v6n3p62>
- Xie, F., & Patten, A. (2025). Epistemic lexical verbs in academic writing: A corpus-based comparative study of L1 English and L1 Chinese writers of English. *Applied Corpus Linguistics*, 5(3), Article 100144. <https://doi.org/10.1016/j.acorp.2025.100144>
- Xu, Z. (2022). Pragmatic functions of evidentiality in diplomatic discourse: Toward a new analytical framework. *Frontiers in Psychology*, 13, Article 1019359. <https://doi.org/10.3389/fpsyg.2022.1019359>