Consortium of Asia-Pacific
Education Universities (CAPEU)
Journal of Education

Vol.2, No.1, July 2026
Available online at: https://ejournal-dppm.upi.edu/capeu

Cross-Linguistic Variation in Modal Expressions: Evidence from English and Uzbek Newspaper Corpora

Muzaffar Obilov[1*]

[1]*University of Business and Science, Uzbekistan*

**Corresponding author: m.obilov92@gmail.com*

Abstract

Modality is a fundamental linguistic category through which speakers express certainty, obligation, and possibility, yet its realization differs substantially across typologically distinct languages, such as English and Uzbek. This study investigates modal expressions in English and Uzbek newspaper discourse through corpus-based analysis. Comparable corpora were compiled from 100 English newspapers (approximately 50 million words) and 55 Uzbek social-political newspapers (approximately 25 million words) published in 2024. The study employed a mixed-methods approach. The English corpus was analysed using AntConc (Version 4.4.1) to extract nine core modal verbs, while the Uzbek corpus focused on analytical modal constructions. The frequency analysis revealed that will and -sa kerak were the most frequent forms in their respective languages. The semantic categorisation further identified five microfields, with probability expressions representing the largest category in both languages (62% of will instances in English and 48% of all modal expressions in Uzbek). The Pragmatic examination demonstrated that English newspapers employ modals primarily for analytical argumentation and epistemic commitment, whereas Uzbek newspapers use them predominantly for persuasive purposes. The genre-based comparison indicated reduced modal frequency in news reports and elevated usage in analytical genres. The findings contribute to cross-linguistic pragmatics, contrastive semantics, and translation studies while offering pedagogical implications for English language education in the Asia-Pacific region through corpus-informed teaching materials, contrastive grammar instruction, translator training, and learners' pragmatic competence in interpreting modality across linguistic and cultural contexts.

Keywords: modality, modal expressions, newspaper discourse, corpus linguistics, English-Uzbek contrastive analysis, semantic microfields, pragmatic functions, frequency analysis

Article history: Received 6 June 2026, accepted 20 August 2026, available online 20 August 2026.

INTRODUCTION

In contemporary linguistic research, modality represents one of the most complex and multifaceted grammatical categories, expressing the speaker's attitude toward propositional content and encompassing fundamental meanings of necessity, possibility, probability, obligation, and capability. The philosophical and logical foundations of modality can be traced back to Aristotle, who established the fundamental distinction between necessary and contingent propositions, conceptualizing modality as the expression of whether objective reality permits or necessitates the actualization of actions or states (Aristotle, ca. 350 BCE/1984; Lyons, 1977). This foundational work initiated a long tradition of modal investigation spanning philosophy, logic, and eventually linguistics.

During the medieval period, the modalist school of thought emerged in Western Europe under Aristotelian influence, making significant contributions to distinguishing between *dictum* (objective propositional content) and *modus* (subjective attitudinal component). This conceptual differentiation laid the essential groundwork for modern linguistic approaches to modality (Bally, 1932). Charles Bally subsequently refined these concepts, proposing that every utterance contains two fundamental components: *dictum* representing objective-informational content and *modus* representing subjective-emotional evaluation (Akhmanova, 1966).

Contemporary theoretical frameworks conceptualize modality as a linguistic category encompassing both the speaker's attitude toward propositional content and the relationship between utterance content and reality (Narrog, 2012; Palmer, 2001). Vinogradov's (1975) influential typology distinguishes between objective modality, which expresses the relationship between propositional content and reality, and subjective modality, which reflects the speaker's personal position, emotions, and perspectives toward reality. This distinction has profoundly influenced Slavic and Turkic linguistic traditions.

ISSN 2443-2555 (online) 2598-6333 (print) © 20XX The Authors. Published by Universitas Pendidikan Indonesia.
This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)
http://doi.org/10.17509/capeu.X.X.36–45

From a functional-communicative perspective, modality serves multiple critical purposes in discourse. First, it expresses the speakers' attitude toward reality and enables speakers to manifest their personal positions alongside factual statements. Second, it reflects the speakers' attitude toward propositional content, expressing dimensions such as truthfulness, necessity, and probability. Third, it conveys emotional states, subjective evaluations, and worldviews through various linguistic resources. Fourth, it performs communicative-pragmatic functions by managing discourse, enhancing persuasive impact, and shaping listener attitudes (Halliday, 1970; Palmer, 2001).

Typologically, modal systems exhibit significant cross-linguistic variation. Palmer's comprehensive framework distinguishes dynamic modality (expressing ability and volition), deontic modality (expressing obligation and permission), and epistemic modality (expressing degrees of certainty regarding propositional truth) (Palmer, 2001). Lyons proposes an alternative dichotomous framework contrasting epistemic and deontic modality as the two fundamental domains (Lyons, 1977). These typological differences manifest distinctly in English and Uzbek modal systems.

English modal verbs constitute a restricted, highly grammaticalized auxiliary class exhibiting distinctive syntactic properties, including defective paradigms, an absence of non-finite forms, and strict co-occurrence restrictions (Jespersen, 1924). While Germanic languages, including English and German, have developed these modal verbs as primary devices for expressing modality, Turkic and Finno-Ugric languages, such as Uzbek, utilize a different typological approach. These languages express modal meanings predominantly through diverse morphological and lexical resources, relying on suffixes, particles, and specialized modal words rather than a distinct class of grammaticalized auxiliary verbs (Johanson & Csató, 1998).

Newspaper discourse presents a particularly rich and significant domain for investigating modal expression patterns. Journalistic texts inherently balance multiple communicative functions: informative objectivity and analytical interpretation, factual reporting and persuasive argumentation, and authoritative assertion and hedged speculation (Bednarek, 2006). The genre-specific demands of newspaper discourse necessitate sophisticated modal strategies for managing epistemic commitment, constructing evidential stance, negotiating social authority, and engaging reader participation (Kotelova, 1978). According to Coates (1983), proficient command of modal resources enables speakers to express their thoughts with clarity, persuasiveness, and credibility; writers can vividly convey their attitudes, emotions, and psychological nuances, while journalists can express opinions clearly and persuade readers of the validity of their positions.

Despite its recognized importance, systematic cross-linguistic comparison of modal expression patterns remains significantly underdeveloped, particularly for typologically distinct language pairs such as English and Uzbek. The primary research objectives are threefold: to conduct a comprehensive frequency analysis of modal expressions in English and Uzbek newspaper corpora, establishing quantitative baselines; to categorize modal expressions according to semantic microfields, revealing language-specific distributions and universal tendencies; and to analyze pragmatic functions and genre-based patterns, illuminating relationships between modal choice and communicative purpose. This study contributes to the theoretical understanding of cross-linguistic modal systems and to practical applications in translation pedagogy, contrastive lexicography, and journalistic writing instruction.

LITERATURE REVIEW

The investigation of modality has evolved through multiple paradigms, from philosophical-logical foundations through structural-grammatical approaches to contemporary functional-pragmatic frameworks. The term *modality* originated in logic, deriving from the French *modal* and the Medieval Latin *modalis*, meaning pertaining to manner or mode. The Latin *modus* encompasses multiple meanings including measure, quantity, proper measure, rhythm, song, way, method, fashion, and style (Bhatia, 1993). Aristotle's work initiated systematic modal investigation, proposing that modality expresses the actualization or potential non-actualization of actions in objective reality, representing core logical concepts that reflect fundamental properties of objective existence (Aristotle, ca. 350 BCE/1984).

Medieval modalists significantly advanced the differentiation between *dictum* and *modus*, interpreting *dictum* as objective propositional content and *modus* as subjective evaluative component (Modistae, 1990). Charles Bally

developed these concepts further, articulating the two fundamental structural components of utterance content: *dictum* as the objective-informational component and *modus* as the subjective-emotional component (Bally, 1932).

Vinogradov's (1975) theoretical framework distinguishes objective modality from subjective modality. Objective modality is expressed through grammatical resources, particularly verbal mood systems, while subjective modality manifests through diverse lexical, grammatical, and prosodic resources. This framework has profoundly influenced Russian and Turkic linguistic scholarship. Furthermore, investigating modality in connection with formal logic and through specific linguistic units remains crucial; the theoretical correspondence between these two fields continues to play a major role in semantic analysis across different languages (Lyons, 1977; Palmer, 2001).

Palmer's cross-linguistic survey established an influential tripartite classification: dynamic modality (expressing the subject's ability or capacity to perform actions), deontic modality (expressing obligation, permission, and prohibition regarding propositional content), and epistemic modality (expressing degrees of the speaker's certainty regarding propositional truth) (Palmer, 2001).

Different linguistic resources realize modal meanings across language types, manifesting through various levels: grammatical, lexical, and prosodic. Grammatical resources, including modal verbs and other morphological forms, represent the most widespread and systematic means for expressing modality. Lexical resources, particularly specialized modal lexemes and expressions, also serve to convey specific modal meanings. Furthermore, prosodic features such as tone, intonation, and stress are highly productive; through prosodic shifts, even structurally declarative sentences can convey imperative or demanding meanings (Depraetere et al., 2023).

Building on the subjective evaluative component of the modus, the scope of modal content and the interaction of its expressive resources are heavily influenced by textual presupposition. This presupposition relies on the speaker's subjective knowledge about reality and their specific level of certainty regarding the phenomena being discussed (Lyons, 1977; Palmer, 2001).

Corpus-based approaches have emerged as particularly influential methodologies in recent decades, enabling systematic quantitative analysis across large-scale datasets (Biber et al., 1999). Corpus-based studies reveal significant genre-based variation: academic discourse favors epistemic hedging modals, conversational discourse prefers deontic modals, and newspaper discourse exhibits a relatively balanced distribution across multiple modal categories (Hunston & Thompson, 2000). Research specifically addressing newspaper discourse identifies modality as crucial for constructing journalistic stance, managing information sources, and engaging reader participation (Bednarek, 2006; Fairclough, 1995). Modal expressions function not merely to convey semantic content but serve complex pragmatic purposes including hedging controversial claims, attributing propositions to external sources, and constructing authorial credibility (Fowler, 1991).

Contrastive modal studies across languages remain relatively limited, with most research focusing on European language pairs (van der Auwera & Plungian, 1998). Although modality has received extensive scholarly attention, most existing accounts have examined it either from a single-language perspective or within broad theoretical frameworks (Coates, 1983; Fraser, 2010; Huddleston & Pullum, 2002; Palmer, 2001). Historically, research on Turkic languages has concentrated on describing individual grammatical systems rather than systematic comparisons with English (Johanson & Csató, 1998). However, recent scholarship has begun to bridge this typological divide. Üzüm (2019) details how epistemic modality in Uzbek relies on distinct syntactic properties and morphological markers rather than grammaticalized auxiliaries. Expanding on this, Mustafayeva (2025) demonstrates that while English depends heavily on a closed class of modal verbs, Uzbek achieves similar pragmatic functions through a diverse range of lexico-grammatical devices, including predicative words like *kerak* and *mumkin*. Furthermore, Sanaqulova (2023) highlights how context shapes the lexico-semantic realization of polysemous modals differently across English and Uzbek narratives. Despite these foundational contributions, few studies have combined corpus-based quantitative analysis with qualitative semantic-pragmatic interpretation to examine how these typological differences function in journalistic contexts (Depraetere et al., 2023). Consequently, empirical evidence comparing modal expressions across English and Uzbek in authentic newspaper discourse remains scarce. This study addresses this gap through comprehensive corpus-based comparison, employing an integrated mixed-methods corpus approach.

METHODS

This study employs a mixed-methods design, combining a quantitative corpus-based comparative methodology with qualitative semantic-pragmatic investigation, to examine modal expressions in English and Uzbek newspaper discourse. The methodological framework draws upon established corpus linguistic techniques while adapting analytical procedures to accommodate typological particularities.

The English newspaper corpus consists of 100 diverse publications published throughout 2024, encompassing broadsheet newspapers (*The Guardian*, *The New York Times*, *The Washington Post*), tabloid publications, specialized business periodicals, and regional newspapers to ensure representative coverage of journalistic styles and readership demographics. The corpus contains approximately 50 million words, providing sufficient statistical power for reliable frequency analysis. The Uzbek newspaper corpus comprises 55 social-political newspapers from 2024, including both national publications (*Xalq so'zi*, *O'zbekiston ovozi*, *Darakchi*) and regional newspapers, totaling approximately 25 million words.

Corpus compilation involved the systematic collection of digital newspaper texts, with careful attention to genre representation including news reports, analytical articles, opinion pieces, editorials, interviews, and feature stories. This genre diversity enables investigation of modal variation across different communicative functions. Text processing included format standardization, removal of non-textual elements, and metadata annotation (publication source, date, genre category, and topic domain) to facilitate subsequent subcorpus analysis (McEnery & Brezina, 2022).

For the English corpus, modal verb extraction employed AntConc (Version 4.4.1) concordance software utilizing specialized search queries to identify nine core modal auxiliaries: *will*, *would*, *can*, *could*, *may*, *might*, *must*, *should*, and *shall* (Anthony, 2026). Search procedures incorporated immediate context windows to filter non-modal homonyms (e.g., the nominal *will* and the verbal *can*) and to capture modal-adjacent elements relevant for semantic analysis. Each identified instance was extracted with sufficient context to enable subsequent manual verification and semantic categorization.

For the Uzbek corpus, modal expression identification focused on primary analytical constructions documented in Uzbek grammatical descriptions: the hypothetical-conditional suffix *-sa* combined with the necessity particle *kerak* (forming *-sa kerak*, expressing probability), the possessive construction with *mumkin* (forming *-(i)shi mumkin*, expressing possibility), the capability auxiliary *-a olmoq* (expressing ability), the necessity constructions *-(i)sh kerak*, *-(i)sh lozim*, *-(i)sh shart*, and *-(i)sh zarur* (expressing varying degrees of obligation and necessity), and additional modal particles *mumkin*, *-dir*, and *-(a)r*. The agglutinative morphology of Uzbek necessitated morpheme-level search patterns to accommodate variant surface realizations.

Frequency normalization employed standard corpus linguistic procedures, calculating relative frequencies per million words using the formula Normalized Frequency = (Raw Frequency / Total Corpus Words) × 1,000,000. This normalization facilitates valid statistical comparison between the English and Uzbek corpora despite their differing absolute sizes.

Semantic categorization involved assigning each modal instance to one of five predefined semantic microfields: probability (expressing the likelihood of a proposition being true), possibility (expressing circumstantial feasibility), capability (expressing an agent's ability), necessity (expressing requirement or essentiality), and obligation (expressing deontic commitment or duty). This categorization drew upon established modal semantic frameworks (Narrog, 2012; Palmer, 2001). For polysemous modals admitting multiple interpretations, context-sensitive analysis determined the predominant reading. Inter-rater reliability for semantic categorization achieved a Cohen's kappa coefficient of 0.87, indicating strong agreement (Landis & Koch, 1977).

Pragmatic functional analysis examined the discourse-level purposes served by modal expressions, including hedging strategies, evidential marking, persuasive argumentation, reader engagement, and stance construction (Fairclough, 2003). In addition, modal frequencies were compared across six primary newspaper genres, and analysis of variance (ANOVA) was used to determine whether the observed differences were statistically significant.

RESULTS

This section presents a comprehensive analysis of modal expression patterns in English and Uzbek newspaper discourse, reporting quantitative frequency distributions, semantic categorizations, pragmatic functions, and genre-based variations.

Frequency Analysis of English Modal Verbs

Analysis of the English newspaper corpus revealed substantial variation in modal usage. Table 1 presents raw frequencies and normalized frequencies (per million words) for nine core English modal auxiliaries.

Table 1
Frequency Distribution of Modal Verbs in the English Newspaper Corpus

Modal verb	Raw frequency	Per million words	Semantic domain
will	2,571	51.42	Future prediction, volition
would	2,031	40.62	Hypothetical, conditional
can	1,749	34.98	Possibility, capability
could	1,255	25.10	Hypothetical possibility
may	779	15.58	Epistemic possibility
should	644	12.88	Obligation, expectation
might	431	8.62	Weak epistemic possibility
must	295	5.90	Strong necessity
shall	12	0.24	Future commitment (formal)
Total	**9,767**	**195.34**	

Table 1 indicates that *will* emerged as the most frequent modal verb, occurring 2,571 times (51.42 instances per million words). This high frequency reflects the central role of future-oriented prediction and speculation in newspaper discourse, particularly in articles addressing political developments, economic forecasts, and policy implications. *Would* appears 2,031 times (40.62 per million words), functioning primarily in conditional constructions, reported speech contexts, and hypothetical reasoning typical of analytical journalism.

The modal verb *can* demonstrates a high frequency (1,749 instances, 34.98 per million words), serving multiple semantic functions including dynamic ability, circumstantial possibility, and deontic permission. Its polysemy contributes to distributional prominence across diverse journalistic contexts. *Could* appears 1,255 times (25.10 per million words), predominantly in conditional and hypothetical constructions where counterfactual reasoning or tentative speculation is required.

The epistemic modals *may* (779 instances, 15.58 per million words) and *might* (431 instances, 8.62 per million words) exhibit moderate frequencies, functioning to hedge assertions and acknowledge uncertainty. The deontic modals *should* (644 instances, 12.88 per million words) and *must* (295 instances, 5.90 per million words) appear with lower frequencies, reflecting constrained use of explicit obligation expressions. *Shall* demonstrates an extremely low frequency (12 instances, 0.24 per million words), restricted almost exclusively to legal and formal contexts.

Semantic Distribution in English Newspapers

Semantic analysis reveals that probability meanings, expressed primarily through *will* (62% of instances), *would* (43%), *may* (73%), and *might* (92%), constitute the largest semantic domain. The high percentage of *might* instances expressing probability (92%) underscores its specialization as a weak epistemic modal. Possibility and capability meanings, predominantly realized through *can* (56% of instances) and *could* (59%), represent the second major semantic domain. Necessity and obligation meanings, expressed through *should* (70% of instances) and *must*

(76%), constitute a smaller but pragmatically significant category concentrated in opinion pieces and editorials (Bhatia, 1993).

Frequency Analysis of Uzbek Modal Expressions

Table 2 presents frequency data for primary Uzbek modal constructions identified in the newspaper corpus. The data reveal that Uzbek employs a more diverse set of analytical modal constructions compared to the restricted inventory of English, reflecting the agglutinative morphological profile of Turkic languages.

Table 2
Frequency Distribution of Modal Expressions in the Uzbek Newspaper Corpus

Modal construction	Raw frequency	Per million words	Semantic domain
-sa kerak	3,240	129.60	Probability, desirability
-(i)shi mumkin	2,156	86.24	Possibility, feasibility
-(i)sh kerak	2,104	84.16	Necessity, obligation
mumkin	1,893	75.72	Possibility, permission
-dir	1,876	75.04	Epistemic certainty
-a olmoq	1,547	61.88	Capability, ability
-(a)r	1,342	53.68	Habitual action
-(i)sh lozim	876	35.04	Necessity, appropriateness
-(i)sh shart	543	21.72	Strong necessity
-(i)sh zarur	423	16.92	Essential necessity
Total	**16,000**	**640.00**	

The most frequent Uzbek modal construction is *-sa kerak* (3,240 instances, 129.60 per million words), combining the conditional suffix *-sa* with the necessity particle *kerak* to express epistemic probability with evaluative overtones. The possibility constructions *-(i)shi mumkin* (2,156 instances, 86.24 per million words) and *mumkin* (1,893 instances, 75.72 per million words) demonstrate substantial frequency, functioning to express circumstantial possibility and feasibility.

Necessity and obligation expressions show diverse realization: *-(i)sh kerak* (2,104 instances, 84.16 per million words), *-(i)sh lozim* (876 instances, 35.04 per million words), *-(i)sh shart* (543 instances, 21.72 per million words), and *-(i)sh zarur* (423 instances, 16.92 per million words). This lexical differentiation enables fine-grained distinctions among degrees of necessity that English expresses primarily through context-dependent interpretation of *must* and *should.*

Semantic Distribution in Uzbek Newspapers

Semantic categorization of Uzbek modal expressions reveals that probability meanings constitute the largest category, representing approximately 48% of total modal usage. This category is primarily realized through *-sa kerak*, *-dir*, and contextual uses of *mumkin* with probability interpretations. Necessity and obligation meanings represent 32% of Uzbek newspaper modal usage, appearing through differentiated constructions; this proportion exceeds the English necessity category, potentially reflecting cultural-rhetorical traditions emphasizing explicit normative evaluation. Possibility and capability meanings account for 20% of modal usage, realized through *-(i)shi mumkin*, *mumkin*, and *-a olmoq* constructions.

Cross-Linguistic Comparison and Genre Analysis

Comparative analysis reveals both convergent and divergent patterns. Both languages demonstrate probability expressions as the most frequent semantic category, confirming that journalistic discourse commonly requires extensive linguistic resources for prediction and speculation. However, specific modal forms reflect

typological constraints: English relies heavily on the grammaticalized auxiliary *will*, while Uzbek distributes probability functions across multiple analytical constructions.

Genre-based analysis reveals parallel patterns: news reports in both languages employ reduced modal density compared to analytical articles, opinion pieces, and editorials. This finding supports the universal principle that factual reporting genres favor assertive declaration over modal qualification, while interpretive genres require modal expressions to hedge claims and acknowledge uncertainty. Statistical analysis (ANOVA) confirmed significant genre effects on modal frequency in both languages (English: $F(5, 94) = 23.47$, $p < .001$; Uzbek: $F(5, 49) = 18.92$, $p < .001$).

Pragmatic functional analysis identifies both shared and language-specific patterns. In both languages, modal expressions serve crucial hedging functions. English modals demonstrate sophisticated evidential functions, marking information source through modal choice. Uzbek modal constructions show a greater orientation toward persuasive and directive functions, with necessity expressions frequently employed to advocate policies or recommend actions.

DISCUSSION

The present findings illuminate fundamental properties of modal expression in newspaper discourse while raising important theoretical questions regarding cross-linguistic comparison of functionally complex grammatical categories.

Overall, the results both confirm and extend prior scholarship on modality. They confirm the centrality of epistemic, probability-oriented modal meaning in newspaper discourse and the influence of genre on modal frequency, findings consistent with earlier corpus-based accounts showing that epistemic modality is a defining characteristic of journalistic language, since news reporting frequently involves prediction, inference, and evaluation rather than absolute certainty (Hyland, 1998; Palmer, 2001; Biber et al., 1999). At the same time, the study extends this literature by showing that, although English and Uzbek converge in favoring probability as the dominant semantic category, the two languages realize this preference through markedly different grammatical resources—a restricted auxiliary system in English interpreted largely through context versus a diverse set of analytical constructions in Uzbek—a finding that complements typological descriptions of Turkic languages (Johanson & Csató, 1998) thereby providing new cross-linguistic corpus evidence for a typologically distant language pair that has received comparatively little empirical attention.

Theoretical Implications

The corpus-based findings confirm that newspaper discourse constitutes an important site for investigating modal semantics and pragmatics. The predominance of probability expressions in both languages supports the view that prediction and speculation represent core communicative functions in news media. The semantic distribution patterns provide evidence for both universal and language-specific aspects of modal meaning. The five-way semantic categorization demonstrates cross-linguistic applicability, suggesting that these categories represent fundamental dimensions recognized across typologically distinct languages.

The finding that Uzbek employs multiple differentiated necessity constructions contrasts with the reliance of English on context-dependent interpretation, exemplifying a fundamental typological difference. The agglutinative morphology of Uzbek enables productive combination of verbal elements with diverse modal particles, creating a richer range of morphological expression of modal nuances. The restricted auxiliary inventory of English compensates through greater reliance on pragmatic inference and contextual interpretation.

Pragmatic Functions and Genre Variation

The pragmatic analysis reveals that modal expressions function as crucial resources for constructing journalistic stance and managing reader engagement. English newspaper modals demonstrate sophisticated evidential functions, with modal choice systematically correlating with information source. Uzbek newspaper modals show stronger orientation toward persuasive and directive functions, which may reflect different cultural-rhetorical traditions regarding journalistic authority and social responsibility.

The finding that modal frequency varies systematically across newspaper genres confirms that genre represents a crucial variable mediating modal usage patterns. News reports demonstrate reduced modal density as journalists strive for assertive presentation of verified information. Analytical articles and opinion pieces employ elevated modal frequency for prediction, speculation, and normative evaluation. This genre-based variation illustrates the principle that linguistic form adapts to communicative function.

Implications for Translation and Pedagogy

The contrastive findings have significant implications for translation theory and practice. The semantic non-equivalence of modal categories across English and Uzbek complicates straightforward translation, as source-language modals often lack single target-language correspondences. Successful translation requires a thorough understanding of modal semantics, pragmatic functions, and genre conventions in both languages.

Pedagogical applications extend to second-language teaching and journalistic writing instruction, particularly within the Asia-Pacific region, where English is widely taught as a second or foreign language. Corpus-derived frequency data and authentic examples can inform syllabus design and data-driven learning materials for language learners. Contrastive analysis alerts instructors to potential cross-linguistic transfer errors and can guide targeted instructional activities that strengthen learners' pragmatic competence in interpreting modality. Journalistic writing, translation, and English language education programs across the region can incorporate modal analysis to cultivate students' stylistic awareness, intercultural communicative competence, and genre competence.

CONCLUSION

This corpus-based investigation of modal expressions in English and Uzbek newspaper discourse has revealed systematic patterns in modal frequency, semantic distribution, pragmatic function, and genre variation, illuminating both similarities and differences in how typologically distinct languages deploy modal resources for journalistic communication.

Quantitative analysis established that both languages employ modal expressions extensively, though with important typological differences. English relies on a restricted set of highly grammaticalized auxiliary verbs, while Uzbek employs diverse analytical constructions combining verbal elements with modal particles and suffixes. Semantic categorization revealed that probability expressions constitute the most frequent modal category in both languages, reflecting journalism's orientation toward prediction and future-oriented analysis.

The distributional differences between languages may reflect typological affordances and cultural-rhetorical traditions. English shows relatively balanced representation across semantic microfields, while Uzbek demonstrates stronger concentration in the probability (48%) and necessity and obligation (32%) domains. Pragmatic analysis identified both universal and language-specific functions: both languages employ modals for epistemic hedging, while English modals demonstrate more elaborate evidential specialization and Uzbek modals show stronger persuasive orientation.

Genre-based analysis confirmed that modal usage varies systematically across newspaper genres in both languages. The consistency of genre effects suggests that genre appears to represent a more fundamental variable than language typology in determining modal frequency patterns. The findings have significant implications for translation theory, requiring sophisticated decision-making regarding semantic and pragmatic features. Pedagogical applications extend to language learning and journalistic writing instruction.

Despite its contributions, this study has limitations. The single-year corpus might not capture diachronic trends; historical comparison could illuminate whether patterns reflect stable characteristics or contemporary trends. Variation across individual newspaper publications also warrants further investigation. More fine-grained semantic analysis employing experimental methods could validate corpus-based categorizations. Expansion to additional language pairs would test the generalizability of findings.

In conclusion, this research demonstrates that corpus-based methodologies provide powerful tools for examining modal expression across languages. While universal communicative demands generate certain parallel patterns, typological structure and cultural-rhetorical traditions systematically shape how languages realize modal meanings. Understanding these patterns enriches both linguistic theory and practical applications, advancing the empirical foundation for cross-linguistic semantics, pragmatics, and discourse analysis. For English language education in the Asia-Pacific region, these findings suggest several practical directions. Corpus-derived frequency and functional profiles can inform the design of teaching materials and data-driven learning tasks that raise learners' awareness of modal meaning. In addition, contrastive analysis can help instructors anticipate transfer-related difficulties between Uzbek and English modal systems, while translator training programs can use the identified functional differences to strengthen learners' pragmatic competence in rendering modality across languages and cultures.

REFERENCES

Akhmanova, O. S. (1966). *Dictionary of linguistic terms*. Soviet Encyclopedia.

Anthony, L. (2026). *AntConc* (Version 4.4.1) [Computer software]. Waseda University.

Aristotle. (1984). Prior analytics. In J. Barnes (Ed.), *The complete works of Aristotle* (Vol. 1, pp. 39–113). Princeton University Press. (Original work published ca. 350 BCE)

Ashley, K. D. (2017). *Artificial intelligence and legal analytics: New tools for law practice in the digital age*. Cambridge University Press. https://doi.org/10.1017/9781316761380

Bally, C. (1932). *Linguistique générale et linguistique française*. Leroux.

Bednarek, M. (2006). *Evaluation in media discourse: Analysis of a newspaper corpus*. Continuum.

Bhatia, V. K. (1993). *Analysing genre: Language use in professional settings*. Longman.

Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Longman.

Coates, J. (1983). *The semantics of the modal auxiliaries*. Croom Helm.

Depraetere, I., Cappelle, B., Hilpert, M., De Cuypere, L., Dehouck, M., Denis, P., Flach, S., Grabar, N., Grandin, C., Hamon, T., Hufeld, C., Leclercq, B., & Schmid, H.-J. (2023). *Models of modals: From pragmatics and corpus linguistics to machine learning*. De Gruyter Mouton.

Fairclough, N. (1995). *Media discourse*. Arnold.

Fairclough, N. (2003). *Analysing discourse: Textual analysis for social research*. Routledge. https://doi.org/10.4324/9780203697078

Fowler, R. (1991). *Language in the news: Discourse and ideology in the press*. Routledge.

Fraser, B. (2010). Pragmatic competence: The case of hedging. In G. Kaltenböck, W. Mihatsch, & S. Schneider (Eds.), *New approaches to hedging* (pp. 15–34). Emerald Group Publishing. https://doi.org/10.1163/9789004253247_003

Halliday, M. A. K. (1970). Functional diversity in language as seen from a consideration of modality and mood in English. *Foundations of Language, 6*(3), 322–361.

Huddleston, R., & Pullum, G. K. (2002). *The Cambridge grammar of the English language*. Cambridge University Press.

Hunston, S., & Thompson, G. (2000). *Evaluation in text: Authorial stance and the construction of discourse*. Oxford University Press.

Hyland, K. (1998). *Hedging in scientific research articles*. John Benjamins.

Jespersen, O. (1924). *The philosophy of grammar*. Allen & Unwin.

Johanson, L., & Csató, É. Á. (Eds.). (1998). *The Turkic languages*. Routledge.

Kotelova, N. Z. (1978). The first experience in describing Russian neologisms. In *New words and dictionaries of new words* (pp. 5–26). Nauka.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174. https://doi.org/10.2307/2529310

Lyons, J. (1977). *Semantics* (Vol. 2). Cambridge University Press.

McEnery, T., & Brezina, V. (2022). *Fundamental principles of corpus linguistics*. Cambridge University Press.

Modistae. (1990). *Grammatica speculativa*. In G. L. Bursill-Hall (Ed.), *Medieval grammatical theory*. John Benjamins. (Original work published 13th century)

Mustafayeva, M. B. (2025). The lexico-grammatical features of modal words in English and Uzbek languages. *American Journal of Alternative Education*, *2*(3), 25–30. https://scientificbulletin.com/index.php/AJAE/article/view/727/969

Narrog, H. (2012). *Modality, subjectivity, and semantic change: A cross-linguistic perspective*. Oxford University Press.

Palmer, F. R. (2001). *Mood and modality* (2nd ed.). Cambridge University Press.

Sanaqulova, M. O. (2023). Lexico-semantic features of modal verbs in fairy-tales (on the example of English and Uzbek fairy-tales). *International Journal of Social Science & Interdisciplinary Research*, *12*, 93–95. https://zenodo.org/records/14780708

Üzüm, M. (2019). Epistemik kiplik işaretleyicilerinin söz dizimsel özellikleri: Özbek ve Türkiye Türkçesi [Syntactic properties of epistemic modality in Turkish and Uzbek]. *Türkoloji Dergisi*, *23*(2). https://www.researchgate.net/publication/350466744

van der Auwera, J., & Plungian, V. A. (1998). Modality's semantic map. *Linguistic Typology*, *2*(1), 79–124. https://doi.org/10.1515/lity.1998.2.1.79

Vinogradov, V. V. (1975). On the category of modality and modal words in Russian language. In *Selected works: Studies in Russian grammar* (pp. 53–87). Nauka.