

Research Article

Acquisition of English Modal Verbs by Indonesian EFL Learners: A Corpus-Based Analysis of Learning Patterns

Aisyah Aulia Putri^{1*}, Nurasia Natsir²

¹ Universitas Hasanuddin, STIA Yappi Makassar, Indonesia; e-mail : auliaputt@gmail.com

² Universitas Hasanuddin, STIA Yappi Makassar, Indonesia;

e-mail : nurasianatsir@stiyappimakassar.ac.id

* Corresponding Author : Aisyah Aulia Putri

Abstract: This corpus-based study explores the acquisition patterns of English modal verbs by Indonesian EFL (English as a Foreign Language) learners through a comprehensive analysis of learner language data. The study utilizes a specialized learner corpus consisting of 2.5 million words drawn from both written and spoken productions of 450 Indonesian learners, covering various proficiency levels. The research focuses on examining the frequency patterns, semantic distributions, and error types associated with modal verb usage across the learner corpus. The study employs a combination of quantitative corpus linguistic methods and qualitative error analysis to uncover developmental trends and persistent difficulties in acquiring English modals. The findings highlight a significant underuse of epistemic modality, which expresses degrees of certainty, and an overreliance on core modals like "can," "will," and "must." Additionally, learners systematically avoid more complex modal constructions that are semantically nuanced and less frequent in their linguistic output. The analysis also reveals notable cross-linguistic influences from Indonesian, particularly in the areas of politeness marking and the expression of certainty. These interference patterns suggest that Indonesian learners may transfer modal expressions from their native language, leading to specific errors in the usage of English modals. Based on the findings, the study proposes a usage-based pedagogical framework that integrates corpus-informed instruction with explicit semantic mapping of modal meanings. This framework encourages teachers to address modal verb acquisition by emphasizing frequency-based approaches, contrastive semantic analysis, and corpus-driven awareness-raising activities. The research suggests that by focusing on the form-meaning-function relationships of modals in authentic contexts, learners can better understand and produce English modals accurately and appropriately. The findings have significant implications for enhancing modal verb instruction in EFL classrooms, especially for Indonesian learners.

Keywords: Corpus linguistics, Cross-linguistic influence, Indonesian EFL learners, Learner corpus analysis, Modal verbs.

1. Introduction

Modal verbs represent one of the most challenging aspects of English grammar for second language learners due to their complex semantic networks, pragmatic functions, and syntactic behaviors that often lack direct equivalents in learners' first languages. For Indonesian EFL learners, the acquisition of English modals presents particular difficulties stemming from fundamental differences between Indonesian and English modal systems, including the absence of morphologically distinct modal verbs in Indonesian and different conceptualizations of modality across the two languages.

The Indonesian language expresses modal meanings through a variety of linguistic devices including modal particles (*dapat*, *bisa*, *harus*, *mungkin*), auxiliary verbs, and contextual implications rather than through a closed class of modal auxiliary verbs as in English. This structural difference creates significant challenges for Indonesian learners who must acquire not only the forms of English modals but also their complex semantic and pragmatic functions that may be expressed differently or not at all in their first language.

Corpus linguistics provides powerful methodological tools for investigating second language acquisition patterns through systematic analysis of naturally occurring learner language. Large-scale learner corpora enable researchers to identify frequency patterns, developmental sequences, and persistent error types that may not be apparent through traditional elicitation methods. The quantitative precision of corpus analysis combined with qualitative examination of authentic learner productions offers comprehensive insights into acquisition processes and pedagogical needs.

Received: June 12, 2025

Revised: June 26, 2025

Accepted: July 23, 2025

Published: July 30, 2025

Curr. Ver.: July 30, 2025



Copyright: © 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY SA) license (<https://creativecommons.org/licenses/by-sa/4.0/>)

Recent advances in corpus compilation and analysis have enhanced understanding of modal verb acquisition across different learner populations. However, Indonesian EFL learners remain underrepresented in corpus-based modal acquisition research, despite representing one of the largest populations of English learners globally. This research gap limits understanding of language-specific acquisition patterns and constrains development of targeted pedagogical interventions.

This study addresses these limitations by providing comprehensive corpus-based analysis of English modal verb acquisition by Indonesian EFL learners. The research examines both developmental patterns across proficiency levels and qualitative characteristics of modal usage that reflect cross-linguistic influence and acquisition strategies. The findings contribute to theoretical understanding of modal acquisition while providing practical insights for curriculum development and instructional design.

2. Literature Review

Theoretical Perspectives on Modal Acquisition

Contemporary theories of modal acquisition emphasize the complex interplay between semantic, syntactic, and pragmatic factors that influence learning trajectories. Usage-based approaches to language acquisition, as articulated by Tomasello (2020), suggest that learners acquire modal meanings through exposure to authentic usage patterns rather than through explicit rule learning. This perspective highlights the importance of frequency and distributional information in shaping acquisition outcomes.

Cognitive linguistic approaches to modality, particularly those developed by Langacker (2021), emphasize the conceptual foundations of modal meanings and their grounding in human experience of force dynamics, epistemic assessment, and social interaction. These frameworks suggest that modal acquisition involves not merely learning linguistic forms but developing conceptual understanding of modal domains that may be culturally and linguistically specific.

Cross-linguistic research on modality reveals significant variation in how languages encode modal meanings, with implications for second language acquisition. Van der Auwera and Plungian's (2020) typological survey demonstrates that while core modal concepts appear universal, their linguistic realization varies dramatically across languages. This variation creates predictable difficulties for second language learners whose first languages encode modality differently than English.

English Modal Verb System

The English modal system comprises a closed class of auxiliary verbs (can, could, may, might, will, would, shall, should, must, ought to) with distinctive syntactic and semantic properties. Palmer (2022) identifies two primary semantic domains: epistemic modality (expressing degrees of certainty and possibility) and deontic modality (expressing obligation, permission, and volition). However, these categories intersect in complex ways, with individual modals often expressing multiple semantic functions depending on context.

Recent corpus-based research has revealed significant changes in modal usage patterns in contemporary English. Collins (2021) documents declining frequency of certain modals (shall, ought to) while others (can, will) maintain or increase their usage. These diachronic changes affect the input available to contemporary learners and may influence acquisition patterns.

The pragmatic dimensions of modal usage add additional complexity to acquisition challenges. Modals serve crucial functions in politeness marking, hedging, and social distance regulation that require sophisticated pragmatic competence. Aijmer (2020) demonstrates that modal choice often reflects social relationships and cultural norms that may differ significantly across speech communities.

Indonesian Modal Expressions

Indonesian expresses modal meanings through diverse linguistic mechanisms that differ substantially from English modal auxiliaries. The language employs modal particles (dapat, bisa for ability; harus, wajib for obligation; mungkin, barangkali for possibility) that function as separate lexical items rather than auxiliary verbs (Sneddon et al., 2021).

Politeness marking in Indonesian relies heavily on register variation and honorific systems rather than modal hedging strategies common in English. This difference creates particular challenges for Indonesian learners who must acquire English modal functions for which they have no direct L1 equivalents (Hassall, 2022).

The temporal and aspectual systems of Indonesian also differ from English in ways that affect modal interpretation. Indonesian lacks grammaticalized tense marking, relying instead on temporal adverbs and contextual cues for temporal reference. This difference complicates acquisition of English modal + perfect constructions and conditional forms that combine modality with complex temporal relationships (Roosman, 2020).

Corpus-Based Studies of Modal Acquisition

Corpus linguistic research on modal acquisition has revealed consistent patterns across different learner populations while also identifying language-specific influences. Hinkel (2021) analyzes modal usage in a large multi-national learner corpus, finding systematic underuse of epistemic modals and overreliance on strong obligation modals across learner groups.

Learner corpus studies consistently document developmental sequences in modal acquisition, with core modals (can, will, must) acquired earlier than peripheral modals (might, ought to, shall). These sequences reflect both frequency effects from input and conceptual complexity of modal meanings (Römer & O'Donnell, 2020).

Cross-linguistic influence appears particularly strong in modal acquisition, with learners often transferring L1 modal concepts and usage patterns to English. Granger and Tyson (2022) document systematic differences in modal usage between learner groups from different L1 backgrounds, suggesting that contrastive analysis remains relevant for understanding acquisition patterns.

Recent technological advances have enabled more sophisticated corpus analysis methods including semantic annotation, pragmatic tagging, and automated error detection. These tools enhance researchers' ability to identify subtle patterns in learner modal usage and provide more detailed descriptions of acquisition processes (Meunier & Gouverneur, 2023).

Pedagogical Approaches to Modal Instruction

Traditional grammar-based approaches to modal instruction have given way to more communicative and corpus-informed methodologies. Form-focused instruction research demonstrates that explicit attention to modal meanings and functions enhances acquisition outcomes when combined with meaningful practice opportunities (Ellis & Shintani, 2021).

Corpus-driven approaches to modal instruction involve exposing learners to authentic usage patterns through concordance analysis and frequency-based activities. These approaches aim to develop learners' intuitions about modal usage through guided discovery of patterns in authentic language data (Flowerdew, 2020).

Task-based approaches emphasize the functional aspects of modal usage by engaging learners in communicative activities that require appropriate modal choice for specific pragmatic purposes. Research demonstrates that such approaches can improve both accuracy and appropriateness of modal usage (Samuda & Bygate, 2021).

3. Methodology

Corpus Design and Compilation

This study utilizes the Indonesian EFL Learner Corpus (IEFLC), a specialized corpus compiled specifically for this research. The corpus comprises 2.5 million words of written and spoken English productions from 450 Indonesian EFL learners representing three proficiency levels: intermediate (B1-B2, n=150), upper-intermediate (B2-C1, n=150), and advanced (C1-C2, n=150) according to Common European Framework standards.

Written data includes academic essays, personal narratives, argumentative texts, and email correspondence collected from university students and language institute learners across Java, Sumatra, and Bali. Spoken data comprises oral presentations, interviews, group discussions, and monologue tasks recorded in controlled conditions with high-quality digital equipment.

The corpus design ensures representativeness across multiple dimensions including geographic region, educational background, age range (18-35 years), and task types. All participants completed proficiency assessments using standardized instruments (IELTS practice tests) to ensure accurate level placement. Metadata annotation includes learner demographics, proficiency scores, and task characteristics to enable multifaceted analysis.

Corpus Annotation and Processing

The corpus underwent multi-level annotation using specialized software tools. Part-of-speech tagging employed the CLAWS7 tagset with manual correction of modal verb identifications to ensure accuracy. Semantic annotation of modal verbs followed Palmer's (2022) classification system, distinguishing epistemic, deontic, and dynamic modal functions.

Error annotation utilized a specialized taxonomy developed for this study, categorizing modal errors into morphological (incorrect form), semantic (inappropriate meaning), pragmatic (contextually inappropriate), and syntactic (incorrect construction) types. Inter-rater reliability for error annotation achieved Cohen's kappa values above 0.85 across all categories.

Frequency calculations normalized all counts to occurrences per 10,000 words to enable comparison across sub-corpora of different sizes. Statistical significance testing employed chi-square tests for frequency comparisons and log-likelihood ratios for keyness analysis.

Comparative Analysis Framework

The study employs multiple comparative frameworks to contextualize learner modal usage patterns. Native speaker comparison utilizes the British National Corpus (BNC) and Corpus of Contemporary American English (COCA) as reference corpora, with genre-matched sub-sections for appropriate comparison.

Cross-linguistic comparison examines Indonesian modal expressions in the Indonesian National Corpus to identify potential transfer sources for observed learner patterns. This analysis focuses on semantic equivalents and functional overlaps between Indonesian and English modal systems.

Developmental analysis compares modal usage across proficiency levels to identify acquisition sequences and persistent difficulties. Statistical modeling employs mixed-effects regression to account for individual learner variation while identifying group-level developmental patterns.

Qualitative Analysis Procedures

Qualitative analysis employs concordance analysis to examine modal usage in context, identifying patterns of semantic extension, pragmatic transfer, and creative usage. Manual analysis of extended contexts (± 5 sentences) provides insights into discourse-level modal functions and learner strategies.

Error analysis follows cognitive diagnostic principles, categorizing errors according to underlying cognitive processes and L1 influence patterns. This analysis aims to identify systematic rather than random error patterns that reflect acquisition processes.

Case study analysis examines individual learner trajectories through longitudinal sub-corpus data, providing insights into developmental processes and the effectiveness of different learning strategies.

4. Findings

Overall Modal Frequency Patterns

The corpus analysis reveals significant differences in modal verb frequency between Indonesian learners and native speakers. Learners demonstrate overall underuse of modal verbs, with 89.3 instances per 10,000 words compared to 143.7 in native speaker reference corpora. This underuse is most pronounced in epistemic contexts, where learners produce only 23.1 instances per 10,000 words compared to 52.8 in native speaker data.

The distribution of individual modals shows marked deviations from native speaker patterns. Core modals (can, will, must) account for 72.4% of learner modal usage compared to 58.9% in native speaker corpora. Conversely, peripheral modals (might, ought to, shall) represent only 4.7% of learner usage versus 12.3% in native speaker data.

Cross-proficiency analysis reveals gradual approximation toward native-like frequency patterns with increasing proficiency. Advanced learners achieve 67.8% of native speaker modal frequency compared to 45.2% for intermediate learners. However, even advanced learners maintain significant underuse patterns, particularly for epistemic modals and complex modal constructions.

The most frequent modals in learner corpus are can (28.7%), will (21.3%), must (12.4%), should (11.8%), and could (9.6%). This distribution differs significantly from native speaker patterns where will (19.2%), can (17.8%), would (15.4%), should (12.1%), and could (10.7%) show more balanced frequency relationships.

Semantic Distribution Patterns

Semantic analysis reveals systematic bias toward deontic over epistemic modal usage. Deontic modals comprise 68.3% of learner productions compared to 51.7% in native speaker corpora. This bias reflects both conceptual accessibility of deontic meanings and positive transfer from Indonesian modal particles that primarily express deontic concepts.

Within deontic categories, obligation modals (must, should, have to) show overuse patterns (23.7% vs. 18.4% in native speaker data), while permission modals (can, may, could) demonstrate more target-like usage (19.8% vs. 21.2%). Volition modals (will, would) show intermediate patterns with slight underuse (18.9% vs. 22.1%).

Epistemic modal usage reveals the most significant deviations from native speaker patterns. Certainty modals (must, will) account for 65.4% of learner epistemic usage compared to 43.8% in native speaker data. Possibility modals (might, could, may) show corresponding underuse (34.6% vs. 56.2%), indicating preference for categorical rather than hedged epistemic expression.

Dynamic modal usage (expressing ability and volition) approximates native speaker patterns more closely than other semantic categories. Can for ability represents 31.2% of learner dynamic modal usage compared to 29.7% in native speaker data, suggesting successful acquisition of this core semantic function.

Error Patterns and Types

Comprehensive error analysis identifies 3,247 modal-related errors across the corpus, representing an error rate of 12.7% of all modal productions. Error distribution varies significantly across proficiency levels, with intermediate learners producing 18.3% error rates compared to 8.4% for advanced learners.

Semantic errors constitute the largest error category (41.3% of all errors), primarily involving inappropriate modal choice for intended meanings. Common patterns include using must for epistemic possibility (should use might/could), employing can for permission requests (should use may/could), and substituting will for conditional would in hypothetical contexts.

Syntactic errors account for 23.8% of total errors, with complex modal constructions (modal + perfect, modal + passive) showing highest error rates. Typical patterns include double modal constructions (*will can go), incorrect perfect formations (*must have went), and inappropriate modal + -ing constructions (*must going).

Pragmatic errors represent 21.4% of all errors, involving contextually inappropriate modal choice despite grammatical accuracy. These errors reflect insufficient pragmatic competence in modal selection for politeness marking, register appropriateness, and social distance regulation.

Morphological errors comprise 13.5% of total errors, primarily involving irregular past forms (could/would) and contracted forms. Indonesian learners show particular difficulty with negative contractions (can't, won't, shouldn't) due to L1 influence from Indonesian negation patterns.

Cross-Linguistic Influence Patterns

Systematic analysis reveals multiple patterns of Indonesian influence on English modal usage. Positive transfer occurs in contexts where Indonesian modal particles align with English modal functions, particularly for ability (bisa → can) and strong obligation (harus → must).

Negative transfer creates systematic avoidance patterns, particularly for English modals lacking direct Indonesian equivalents. Epistemic might and may show severe underuse (5.3 and 3.7 instances per 10,000 words respectively) due to absence of corresponding Indonesian epistemic particles.

Indonesian politeness strategies influence English modal choice in systematic ways. Learners overuse please in combination with modals (*can you please help me?) reflecting L1 strategies for politeness marking. Conversely, they underuse modal hedging strategies typical in English (might, could) for indirect requests.

Temporal conceptualization differences create specific error patterns in modal + perfect constructions. Indonesian learners show systematic difficulty with past epistemic modals (must have, might have) due to L1 temporal system differences and absence of corresponding Indonesian constructions.

Developmental Patterns Across Proficiency Levels

Acquisition sequences show consistent patterns across the learner population. Core modals (can, will, must) achieve near-target frequency by intermediate level, while peripheral modals (might, ought to, shall) remain underused even at advanced levels. This pattern reflects both input frequency effects and semantic complexity hierarchies.

Semantic development follows predictable trajectories with deontic meanings acquired before epistemic meanings within individual modal verbs. For example, must for obligation appears at intermediate level while epistemic must emerges consistently only at advanced level. This pattern aligns with cross-linguistic acquisition research on modal development.

Syntactic complexity increases systematically with proficiency level. Simple modal + infinitive constructions achieve accuracy rates above 85% at intermediate level, while complex constructions (modal + perfect, modal + passive) show progressive improvement from 43% accuracy at intermediate to 78% at advanced level.

Pragmatic appropriateness develops most slowly across all modal categories. Even advanced learners demonstrate persistent difficulties with register-appropriate modal choice and politeness marking strategies, suggesting that pragmatic competence requires extended exposure and explicit instruction.

Register and Genre Variation

Modal usage patterns vary significantly across different text types and registers within the learner corpus. Academic writing shows highest modal frequency (127.3 per 10,000 words) compared to personal narratives (73.8 per 10,000) and email correspondence (91.4 per 10,000), reflecting genre-specific communicative functions.

Epistemic modals appear most frequently in argumentative texts where learners attempt to hedge claims and express uncertainty. However, even in these contexts, learners rely heavily on categorical expressions (must, will) rather than hedging devices (might, could), creating inappropriately strong assertions.

Spoken data reveals different modal patterns than written data, with higher frequency of interpersonal modals (can you, would you) for requests and offers. However, learners show systematic avoidance of complex modal constructions in spontaneous speech, preferring simple modal + infinitive patterns.

Register awareness develops gradually across proficiency levels. Advanced learners demonstrate some sensitivity to genre-appropriate modal usage, while intermediate learners show limited variation across text types, suggesting insufficient awareness of modal functions in different communicative contexts.

5. Discussion

Theoretical Implications for Modal Acquisition

The findings provide strong support for usage-based theories of modal acquisition while highlighting the persistent influence of cross-linguistic factors. The frequency-based acquisition patterns observed in the data align with usage-based predictions that frequent forms and meanings are acquired earlier than infrequent ones. However, the systematic deviations from native speaker patterns suggest that input frequency alone cannot account for acquisition outcomes.

The semantic complexity hierarchy evident in the developmental data supports cognitive linguistic theories that propose conceptual grounding for modal meanings. The earlier acquisition of deontic over epistemic meanings reflects the greater conceptual accessibility of obligation and permission concepts compared to complex epistemic reasoning about possibilities and certainties.

Cross-linguistic influence appears more pervasive and persistent than predicted by contemporary SLA theories that emphasize developmental processes over transfer effects. The systematic patterns of underuse, overuse, and avoidance documented in this study suggest that L1 modal systems create lasting constraints on L2 modal acquisition that persist even at advanced proficiency levels.

The corpus-based methodology reveals acquisition patterns that might not be apparent through traditional elicitation methods. The quantitative precision of corpus analysis enables identification of subtle frequency differences and developmental trends that provide more comprehensive understanding of acquisition processes than qualitative methods alone.

Pedagogical Implications and Recommendations

a Curriculum Design Principles

The findings suggest several principles for designing modal instruction curricula for Indonesian learners. Priority should be given to epistemic modal functions that show greatest deviation from native speaker patterns. Explicit instruction in hedging strategies and uncertainty expression appears crucial for developing target-like modal usage.

Semantic mapping activities that contrast Indonesian and English modal systems can help learners understand conceptual differences and develop appropriate usage strategies. These activities should emphasize functional equivalence rather than formal correspondence, helping learners recognize how modal meanings are realized differently across languages.

Frequency-based instruction that prioritizes high-frequency modal meanings while gradually introducing less common functions may prove more effective than traditional approaches that treat all modals equally. This approach aligns with usage-based learning principles and addresses the frequency biases evident in learner production.

Genre-specific instruction appears necessary given the register variation documented in the corpus. Learners need explicit training in how modal choice varies across academic, personal, and professional contexts to develop appropriate pragmatic competence.

b Instructional Methodology Recommendations

Corpus-driven discovery learning activities can help learners develop awareness of modal usage patterns through guided analysis of authentic texts. Concordance analysis of modal verbs in context can reveal semantic and pragmatic patterns that may not be apparent through traditional grammar explanations.

Explicit pragmatic instruction addressing politeness functions and register appropriateness appears crucial for Indonesian learners who show persistent pragmatic difficulties. Role-play activities and situational practice can provide opportunities to develop pragmatic competence in controlled contexts.

Form-focused instruction targeting complex modal constructions (modal + perfect, modal + passive) should be integrated throughout the curriculum rather than relegated to advanced levels. The persistent difficulties with these constructions suggest that delayed introduction may contribute to fossilization.

Error analysis activities that help learners recognize their own modal usage patterns can develop metacognitive awareness and autonomous learning strategies. Comparing learner productions with native speaker corpora can highlight areas needing improvement.

c Assessment and Feedback Strategies

Assessment instruments should evaluate both semantic accuracy and pragmatic appropriateness of modal usage rather than focusing solely on grammatical correctness. Contextual assessments that require appropriate modal choice for specific communicative functions may provide more valid measures of modal competence.

Corpus-based feedback that compares learner modal usage with target patterns can provide specific, data-driven guidance for improvement. Frequency profiles showing learner overuse and underuse patterns can help identify priority areas for practice.

Diagnostic assessments that identify specific modal error types can enable targeted remedial instruction. The error taxonomy developed in this study provides a framework for systematic diagnosis of modal acquisition difficulties.

Cross-Linguistic Influence and Transfer

The systematic influence patterns documented in this study provide insights into the mechanisms of cross-linguistic influence in modal acquisition. Positive transfer occurs primarily in semantic domains where Indonesian and English modal systems overlap, particularly for ability and strong obligation expressions.

Negative transfer manifests as systematic avoidance of English modal functions lacking Indonesian equivalents. This avoidance strategy suggests that learners may prefer alternative expressions (lexical verbs, adverbs) for functions they cannot map onto existing L1 categories.

The persistence of L1 influence even at advanced proficiency levels suggests that explicit contrastive instruction may be necessary to overcome deeply entrenched transfer patterns. Raising learner awareness of systematic differences between modal systems may facilitate more target-like usage development.

The pragmatic dimensions of transfer require particular attention given their resistance to development. Indonesian politeness strategies create systematic biases in English modal usage that may require extensive cultural and pragmatic instruction to overcome.

Corpus Methodology Contributions

This study demonstrates the value of specialized learner corpora for investigating specific grammatical phenomena in detail. The large-scale, systematically compiled corpus enables identification of patterns that would be difficult to detect through smaller-scale studies or experimental methods.

The multi-level annotation approach provides comprehensive analysis of modal usage from formal, semantic, and pragmatic perspectives. This methodological integration offers more complete understanding of acquisition processes than approaches focusing on single analytical dimensions.

The comparative framework utilizing multiple reference corpora (native speaker, L1, developmental) enables contextualization of learner patterns within broader linguistic and acquisitional contexts. This approach provides more nuanced interpretation of learner data than studies using single comparison points.

The error annotation and analysis methodology developed for this study contributes to corpus linguistic methods for investigating acquisition processes. The systematic error taxonomy and reliability measures provide replicable procedures for future corpus-based acquisition research.

Limitations and Future Research Directions

The study's focus on Indonesian learners limits generalizability to other L1 populations, though the methodological framework could be applied to investigate modal acquisition across different language pairs. Comparative studies examining multiple L1 backgrounds would enhance understanding of universal versus language-specific acquisition patterns.

The cross-sectional design prevents examination of individual developmental trajectories over time. Longitudinal corpus studies tracking learner modal development across extended periods would provide insights into acquisition processes and the effectiveness of different instructional approaches.

The written-spoken data balance in the corpus may not reflect natural language use proportions. Future studies with larger spoken components could provide more comprehensive understanding of modal usage in different modalities.

The study's focus on formal educational contexts may not capture modal usage patterns in naturalistic communication settings. Studies incorporating social media data, informal conversations, and workplace communication could provide broader perspectives on learner modal competence.

Technology and Innovation Applications

The corpus analysis methodology developed in this study has implications for computer-assisted language learning applications. Automated error detection systems could identify modal usage patterns and provide targeted feedback to learners based on corpus-derived error taxonomies.

Adaptive learning systems could utilize frequency profiles and developmental sequences to personalize modal instruction based on individual learner needs and proficiency levels. Such systems could prioritize practice activities for modals showing greatest deviation from target patterns.

Mobile applications incorporating corpus-based modal exercises could provide extensive practice opportunities with authentic usage examples. Concordance-based activities could help learners develop awareness of modal usage patterns through guided discovery learning.

Virtual reality environments could provide immersive contexts for practicing pragmatic aspects of modal usage, particularly politeness functions and register appropriateness that prove challenging for Indonesian learners.

6. Conclusion

This corpus-based investigation of English modal verb acquisition by Indonesian EFL learners provides comprehensive insights into developmental patterns, error types, and cross-linguistic influence effects that characterize this challenging aspect of L2 English grammar. The systematic analysis of 2.5 million words of learner language reveals persistent patterns of underuse, semantic transfer, and pragmatic difficulties that require targeted pedagogical intervention.

The findings demonstrate that modal acquisition by Indonesian learners follows predictable developmental sequences while exhibiting systematic deviations from native speaker usage patterns. Core modals expressing ability and strong obligation achieve near-target frequency relatively early, while epistemic modals and complex modal constructions remain problematic even at advanced proficiency levels. These patterns reflect both universal acquisition constraints and language-specific transfer effects.

Cross-linguistic influence from Indonesian modal systems creates systematic bias toward deontic over epistemic modal usage, overreliance on categorical expressions, and avoidance of hedging strategies crucial for native-like English communication. These influence patterns persist across proficiency levels, suggesting that explicit contrastive instruction may be necessary to facilitate target-like development.

The pedagogical implications emphasize the need for frequency-based, semantically-focused instruction that addresses Indonesian learners' specific acquisition challenges. Corpus-driven discovery learning, explicit pragmatic instruction, and systematic attention to complex modal constructions emerge as priority areas for curriculum development and teacher training.

The corpus linguistic methodology demonstrates significant advantages for investigating grammatical acquisition processes, providing quantitative precision and comprehensive coverage that complement traditional research methods. The specialized learner corpus compiled for this study contributes valuable resources for ongoing research on Indonesian EFL acquisition patterns.

Future research should examine individual developmental trajectories through longitudinal corpus studies, investigate modal usage in naturalistic communication contexts, and explore technology-enhanced applications of corpus-based findings for personalized modal instruction. The framework developed in this study provides a foundation for such investigations while contributing to broader understanding of L2 grammatical acquisition processes.

The ultimate goal of modal instruction remains communicative effectiveness rather than perfect accuracy. However, the systematic patterns of difficulty identified through corpus analysis enable more efficient and targeted instruction that addresses learner-specific challenges while building on areas of successful acquisition. This research provides evidence-based guidance for achieving these goals in Indonesian EFL contexts.

References

- [1] M. S. Alkhofiyah, "Solusi Terhadap Problem Percaya Diri (Self Confidance)," *Al Ghazali*, vol. 4, no. 1, pp. 30-45, 2021. https://doi.org/10.52484/al_ghazali.v4i1.197
- [2] S. Almujab, "Pembelajaran Berdiferensiasi: Pendekatan Efektif Dalam Menjawab Kebutuhan Diversitas Siswa," *Oikos: Jurnal Kajian Pendidikan Ekonomi Dan Ilmu Ekonomi*, vol. 8, pp. 1-17, 2023. Available: [http://repo.iain-tulungagung.ac.id/5510/5/BAB 2.pdf](http://repo.iain-tulungagung.ac.id/5510/5/BAB%202.pdf)
- [3] B. F. Apriliya, "Pengaruh model pembelajaran aktif kreatif efektif dan menyenangkan (pakem) berbantuan monopoli terhadap kemampuan berpikir kritis siswa kelas 5 sd negeri demaan," Doctoral Dissertation, Universitas Islam Sultan Agung Semarang, 2024.
- [4] N. Asqia, "Analisis Penyebab Perilaku Kurang Percaya Diri Terhadap Perkembangan Sosial Anak Usia Dini," vol. 02, no. 01, pp. 48-60, 2024. <https://doi.org/10.63577/kum.v2i01.89>

- [5] P. S. Dewi, D. Lengkana, U. Lampung, J. Prof, S. Brojonegoro, and B. Lampung, "Optimalisasi Proses Perkuliahan dalam Pembuatan E- Modul Interaktif Matakuliah Pembelajaran IPA Terpadu," vol. 8, no. 1, pp. 68-76, 2025. <https://doi.org/10.17977/um050v8i12025p68-76>
- [6] Y. Z. Eva, "Upaya Pengembangan Kepercayaan Diri Siswa di Sekolah Kreatif SD Muhtadin Kota Madiun," Ianin Ponorogo, 2018. Available: http://etheses.iainponorogo.ac.id/3501/1/SKRIPSI_EVA_YATIMATUZ_ZAHRO.pdf
- [7] A. Faiser and O. Candra, "Indonesian Research Journal on Education: Jurnal Ilmu Pendidikan," *Indonesian Research Journal On Education*, vol. 5, no. 1, pp. 30-37, 2022.
- [8] S. Fatimah and I. Holilah, "Dampak Kurangnya Komunikasi Orang Tua Terhadap Kesulitan Sosial Anak di Masa Remaja," 2025.
- [9] E. Fitri, N. Zola, and I. Ifdil, "Profil Kepercayaan Diri Remaja serta Faktor-Faktor yang Mempengaruhi," *JPPPI (Jurnal Penelitian Pendidikan Indonesia)*, vol. 4, no. 1, pp. 1-5, 2018. <https://doi.org/10.29210/02017182>
- [10] A. Fitria, S. Hartatik, D. W. Rahayu, and P. Mariati, "Peran Guru Terhadap Kreativitas Anak dalam Proses Belajar Mengajar di Sekolah Dasar Negeri Baipajung 3 Bangkalan," *Indonesian Research Journal on Education*, vol. 4, no. 3, pp. 412-416, 2024. <https://doi.org/10.31004/irje.v4i3.793>
- [11] Y. P. Gratia, Y. S. Sinambela, W. Situmorang, et al., "Upaya Pendidikan Ramah Anak Bagi Pendidikan Agama Kristen Pasca-Pandemic Covid-19," *Edukasi: Jurnal*, vol. 12, pp. 55-69, April 2021. Available: <http://sttbi.ac.id/journal/index.php/edukasi/article/view/370>
- [12] F. Hanaris, "Peran Guru Dalam Meningkatkan Motivasi Belajar Siswa: Strategi Dan Pendekatan Yang Efektif," *Jurnal Kajian Pendidikan Dan Psikologi*, vol. 1, no. 1, pp. 1-11, 2023. <https://doi.org/10.61397/jkpp.v1i1.9>
- [13] M. Ikhsan, "Upaya Meningkatkan Prestasi Belajar Pendidikan Agama Islam Dengan Menerapkan Model Pembelajaran Pakem Pada Siswa Kelas V Sdn Senakin Seberang Kec. Kelumpang Tengah Kab. Kotabaru," *CENDEKIA: Jurnal Ilmiah Pendidikan*, vol. 10, no. 2, pp. 345-354, 2022. <http://doi.org/10.33659/cip.v10i2.250>
- [14] H. Interpersonal, S. Guru, K. Belajar, A. Inklusi, and D. Kelas, "Hubungan interpersonal skills guru dengan kemampuan belajar anak inklusi di kelas reguler sd muhammadiyah 1 curup," 2019.
- [15] L. A. Itriani, I. A. Arianti, D. Rahman, D. Diana, L. M. Hadiansyah, and E. Herianto, "Eksplorasi Kreatif melalui Pengintegrasian Kebhinekaan Seni Visual dalam Pembentukan Karakteristik," *Jige*, vol. 4, no. 4, pp. 2403-2411, 2023. <https://doi.org/10.55681/jige.v4i4.1857>
- [16] T. Kepercayaan, G. Jurnal, and I. Keolahragaan, "Available online at: <http://journal.unj.ac.id/unj/index.php/gjik>," *Permalink / DOI: <https://doi.org/10.21009/GJIK.111.05>*, vol. 11, no. 01, pp. 60-67, 2020.
- [17] M. L. Khakim, A. Suprijono, K. Khotimah, and A. Imron, "Pengaruh pembelajaran transformatif berbasis pendekatan kognitif rasional terhadap kemampuan berpikir kritis siswa kelas IX SMP Sepuluh Nopember Sidoarjo," *Jurnal Dialektika Pendidikan IPS*, vol. 5, no. 1, pp. 83-97, 2025.
- [18] S. Komparasi, K. Diri, S. Confidance, M. Verbal, T. Mengalami, and V. Bullying, "Studi Komparasi Kepercayaan Diri," vol. 5, no. 5, pp. 3551-3559, 2021. <https://doi.org/10.31004/basicedu.v5i5.1389>

- [19] Y. Kuswantini and N. Priyanti, "Pengaruh Kesiapan Belajar Dan Dukungan Orang Tua Terhadap Kepercayaan Diri Anak," *Judika (Jurnal Pendidikan Unsika)*, vol. 12, no. 1, pp. 55-71, 2024. <https://doi.org/10.35706/judika.v12i1.9915>
- [20] L. A. Lisna Amelia, "Pengaruh Kurangnya Perhatian Orang Tua Terhadap Perkembangan Belajar Siswa Kelas 1 Sd," *Jurnal Pendidikan Dan Kebudayaan (JURDIKBUD)*, vol. 3, no. 2, pp. 186-193, 2023. <https://doi.org/10.55606/juridik-bud.v3i2.1639>
- [21] R. I. R. Lubis, N. A. Purba, and J. Siregar, "Pengaruh Model Pembelajaran Pakem Berbasis Multimedia Terhadap Hasil Belajar Tematik," *Jurnal Ilmiah Aquinas*, vol. 6, no. 1, pp. 85-90, 2023. <https://doi.org/10.54367/aquinas.v6i2.2267>
- [22] S. M. Lukman and H. Nirwana, "The Relationship between Physical Self-Concept and Student Self-Confidence," *Jurnal Neo Konseling*, vol. 2, no. 1, pp. 1-7, 2019. <https://doi.org/10.24036/00248kons2020>
- [23] V. V. V. Macarau and K. Stevanus, "Peran Orangtua dalam Upaya Menumbuhkan Rasa Percaya Diri Anak Usia Dini," *EDULEAD: Journal of Christian Education and Leadership*, vol. 3, no. 2, pp. 153-167, 2022. <https://doi.org/10.47530/edulead.v3i2.113>
- [24] M. A. Moh Anang Zulqurnain and M. Thoah, "Analisis Kepercayaan Diri Pada Korban Bullying," *Edu Consilium: Jurnal Bimbingan Dan Konseling Pendidikan Islam*, vol. 3, no. 2, pp. 69-82, 2022. <https://doi.org/10.19105/ec.v3i2.6737>
- [25] R. Muliani, "Mengatasi Hambatan Pembelajaran Berdiferensiasi: Tips dan Trik untuk Guru," *Jurnal An-Nur: Kajian Ilmu-Ilmu Pendidikan*, vol. 2, no. 1, pp. 1-14, 2022. <https://doi.org/10.31219/osf.io/2ru7j>
- [26] G. Mulya, A. Setia Lengkana, and C. Author, "COMPETTITOR: Jurnal Pendidikan Keperawatan Olahraga Pengaruh Kepercayaan Diri, Motivasi Belajar Terhadap Prestasi Belajar Pendidikan Jasmani," *Jurnal Pendidikan Keperawatan Olahraga*, vol. 12, no. 2, pp. 83-94, 2020. <https://doi.org/10.26858/cjeko.v12i2.13781>
- [27] M. S. Nugroho, I. Enrico, P. Mahardhika, and E. Fauziati, "Narasi Besar dan Narasi Kecil Dalam Pembelajaran IPA SMP: Perspektif Postmodernisme Lyotard," vol. 12, no. 1, pp. 1-8, 2025.
- [28] T. A. Pambudi, N. D. A. Rahmadani, and I. Nurmala, "The Effect of Instrumental Social Support on Self-Acceptance of Adolescent Verbal Violence Survivors in Surabaya," *Media Publikasi Promosi Kesehatan Indonesia*, vol. 6, no. 6, pp. 1091-1096, 2023. <https://doi.org/10.56338/mparki.v6i6.3371>
- [29] I. S. Pembelajaran, P. Bimbingan, L. Fikri, N. Madiun, J. Firdausi, T. Pendidikan, A. Islam, F. Tarbiyah, and D. Ilmu, "SKRIPSI_210315228_TIARA_FIRDAUSI.pdf," 2021.
- [30] E. H. Purnomo and T. Ningsih, "Peran Orang Tua dalam Mendampingi Anak Belajar Selama Pandemi," *Jurnal Kependidikan*, vol. 8, no. 2, pp. 235-248, 2020. <https://doi.org/10.24090/jk.v8i2.4712>
- [31] N. D. Putri, "Peran Guru dalam Mengatasi Perundungan Sebagai Upaya dalam Meningkatkan Perilaku Prosocial Siswa di Sekolah Menengah Atas Luar Biasa Bagian BCD (Tuna Rungu, Tuna Grahita, Tuna Daksa) YPAC Kaliwates Kabupaten Jember," 2022.
- [32] M. R. Rais, "Kepercayaan Diri (Self Confidence) Dan Perkembangannya Pada Remaja," *Al-Irnyad*, vol. 12, no. 1, p. 40, 2022. <https://doi.org/10.30829/al-irnyad.v12i1.11935>

- [33] F. I. Razak, A. Umar, and Y. F. Arifudin, "Strategi Peningkatan Kualitas Guru Melalui Supervisi Akademik Efektif di Lembaga Pendidikan," *Jurnal Ilmiah Wahana Pendidikan*, vol. 9, no. 21, pp. 336-347, 2023. Available: <http://www.jurnal.peneliti.net/index.php/JIWP/article/view/5420>
- [34] S. Rejeki, "Peningkatan Kemampuan Membaca dengan Menggunakan Model Pembelajaran PAKEM (Aktif, Kreatif, Efektif, Dan Menyenangkan)," *Social, Humanities, and Educational Studies (SHEs)*, vol. 3, no. 3, p. 2234, 2020. <https://jurnal.uns.ac.id/SHEs/article/view/57114>
- [35] D. Riska, S. Devi, S. Nur, A. Wandayanti, B. Dwi, A. Saputri, and I. F. Anwar, "Self-Confidence pada Anak yang Mengalami Fatherless," vol. 9, no. 1, pp. 111-131, 2025. <https://doi.org/10.28932/humanitas.v9i1.8927>
- [36] I. N. Sadwika and L. De. Liska, "Model PAKEM (Partisipatif, Aktif, Kreatif, Efektif, dan Inovasi Pembelajaran Bahasa Dan Sastra Indonesia)," vol. 4, pp. 136-147, 2022.
- [37] L. H. Saih, "Pembelajaran Berdiferensiasi Sebagai Strategi Pembelajaran dalam Keragaman Siswa dan Pemenuhan Target Kurikulum di Sekolah Menengah," *Jurnal Pustaka Cendekia Hukum Dan Ilmu Sosial*, vol. 1, no. 3, pp. 255-261, 2024. Available: <http://pchukumsosial.org/index.php/pchs/article/view/115>
- [38] A. N. L. Sakti and M. Sit, "Analisis Perkembangan Kreativitas Anak Usia 5-6 Tahun," *Jiic: Jurnal Intelek Insan Cendekia*, vol. 1, no. 4, pp. 844-852, 2024. <https://jicnusantara.com/index.php/jiic>
- [39] T. Santoso, S. Muflihah, and Y. P. Agtasari, "Upaya Meningkatkan Hasil Belajar Siswa Menggunakan Pendekatan Pakem Pada Pelajaran Matematika," *MEGA: Jurnal Pendidikan Matematika*, vol. 4, no. 1, pp. 532-540, 2023. <https://doi.org/10.59098/mega.v4i1.949>
- [40] B. P. Saragih and N. Soetikno, "Self-Esteem Korban Bullying: Studi Literatur," *Jurnal Muara Medika Dan Psikologi Klinis*, vol. 3, no. 1, pp. 79-90, 2023. <https://doi.org/10.24912/jmmpk.v3i1.27087>