Research on Mining Development Patterns of Pragmatic Competence in English Learners Assisted by Machine Learning

Yuxia Song*

Liaoning University of International Business and Economics, Dalian, 116052, China
**Corresponding author: 18940929178@163.com*

Abstract: *With the growing demand for cross-cultural communication in the process of globalization, the cultivation of pragmatic competence in English learners has become a critical component of language education. Traditional pragmatics research, which predominantly relies on manual annotation and small-sample analysis, struggles to systematically reveal the developmental trajectories of pragmatic competence. This study, situated at the intersection of pragmatics theory and computational linguistics, conducts research on mining development patterns of pragmatic competence in English learners assisted by machine learning. By constructing a theoretical framework encompassing three dimensions—linguistic pragmatics, social pragmatics, and cognitive processing—the research systematically analyzes explicit and implicit pragmatic features. Subsequently, techniques such as fine-tuning pre-trained language models, multi-modal fusion, and graph neural networks are employed to achieve automated extraction and pattern recognition of pragmatic features. Finally, through multi-dimensional validation and empirical teaching research, the scientific validity and application value of the identified developmental patterns are confirmed. This study offers a new perspective for understanding the mechanisms of pragmatic competence development and provides methodological support for achieving precision in pragmatic instruction.*

Keywords: *pragmatic competence; machine learning; developmental patterns; feature extraction; multi-modal analysis; instructional intervention*

Introduction

Research on the development of pragmatic competence in English learners represents a significant subject within the fields of applied linguistics and second language acquisition. Its central focus lies in revealing how learners progressively acquire the ability to appropriately comprehend and produce discourse within specific socio-cultural contexts. Although traditional research paradigms have constructed numerous theoretical frameworks, they exhibit notable methodological limitations: reliance on intuitive manual judgments, small-scale data analysis, and cross-sectional research designs make it difficult to objectively and systematically capture the dynamic processes and complex patterns of pragmatic competence development. With the large-scale accumulation of learner corpora in the process of educational informatization and the rapid advancement of artificial intelligence technologies, utilizing machine learning methods to mine developmental patterns of pragmatic competence has not only become technically feasible but also emerged as an essential requirement for driving a paradigm shift in this research domain. By integrating pragmatics theory with machine learning technology, this study aims to establish a comprehensive research framework spanning theoretical construction, feature extraction, pattern recognition, and pedagogical application. The necessity of this research lies in overcoming the methodological constraints of traditional approaches, providing a data-driven empirical basis for understanding the intrinsic mechanisms of pragmatic competence development, and offering scientific support for achieving personalized pragmatic instruction, thereby advancing pragmatic competence research toward greater precision and systematicity.

1. Theoretical Construction and Feature Analysis of Pragmatic Competence Development Patterns

1.1 Review and Integration of Theoretical Frameworks for Pragmatic Competence

Research on pragmatic competence is rooted in linguistic philosophy and communication theory, and its conceptual evolution has progressed from a knowledge-based view to a behavioral view, and further to a developmental perspective. Early theories regarded pragmatic competence as an implicit knowledge system about language use rules, focusing on learners' static understanding of concepts such as speech acts and conversational implicatures. Subsequent communicative competence models positioned it within broader socio-cultural contexts, emphasizing its dynamic function in achieving effective communication.

This study integrates these theoretical perspectives to propose a multidimensional operational definition of pragmatic competence: it is an individual's ability, within specific socio-cultural contexts, to comprehensively utilize linguistic resources, contextual cues, and social cognition to appropriately comprehend and produce discourse for achieving communicative intentions. This framework not only encompasses traditional research foci such as the production and comprehension of speech acts and the application of politeness strategies, but also incorporates the construction of discourse coherence and contextual adaptability into its core domains, thereby establishing a theoretical foundation for developing developmental patterns. This integrated framework transcends the limitations of singular theoretical perspectives and provides a conceptual map for characterizing the complex developmental trajectory of pragmatic competence from surface-level formal manipulation to deep socio-pragmatic cognition[1].

1.2 Core Dimensions and Indicator Establishment for Developmental Patterns

Based on the integrated theoretical framework, the developmental patterns of pragmatic competence in English learners can be deconstructed into three interrelated core dimensions: the linguistic pragmatics dimension, the socio-pragmatics dimension, and the cognitive processing dimension. The linguistic pragmatics dimension focuses on the linguistic forms employed to achieve specific communicative functions, with observable indicators including the diversity of speech act strategies, the accuracy and richness of formulaic expressions, and the appropriateness of modality usage. The socio-pragmatics dimension involves sensitivity and adaptability to social conventions and cultural norms within communicative contexts, specifically manifested through considerations of contextual factors such as power and social distance, the selection of politeness levels, and the control of cross-cultural pragmatic transfer. The cognitive processing dimension reflects the efficiency of pragmatic information processing during real-time communication, which can be indirectly inferred through indicators such as hesitations in discourse production, frequency of self-repairs, and response speed to implied meanings.

These three dimensions collectively constitute a dynamic developmental system, wherein linguistic forms serve as the carrier, social cognition functions as the regulatory mechanism, and processing efficiency acts as the supporting condition. The construction of developmental patterns entails a systematic description of the phased characteristics exhibited by learners across each dimension, their developmental trajectories, and the interactive relationships among these dimensions.

1.3 Systematic Classification and Analysis of Pragmatic Features

To achieve quantitative characterization of the aforementioned developmental dimensions, theoretical indicators must be transformed into operational and computable pragmatic features. This study systematically categorizes pragmatic features into two primary classes: explicit features and implicit features. Explicit features primarily refer to elements that can be directly extracted from surface linguistic structures, including lexical-pragmatic features such as the density and distribution of pragmatic markers like mitigators and intensifiers; syntactic-pragmatic features, such as the syntactic realization methods of interrogative and imperative sentences in performing speech acts; and discourse organizational features, such as the use of backchannels and discourse markers in turn-taking[2].

Implicit features involve deeper information that requires inference and computation to obtain, mainly comprising speech act sequence patterns, which refer to the combination and sequential regularities of different speech acts in dialogue; emotion and stance expression features, such as the

emotional polarity, epistemic stance, and attitudinal tendencies conveyed in discourse; and contextual adaptation degree, which indicates the alignment level between linguistic choices and specific contextual parameters. This systematic classification and analysis establishes a methodological foundation for the subsequent application of machine learning techniques to automatically extract and quantify these features from large-scale learner corpora, thereby enabling the discovery and presentation of pragmatic competence development patterns hidden within complex data.

2. Machine Learning-Driven Pragmatic Feature Extraction and Pattern Recognition

2.1 Machine Learning-Based Automated Extraction Methods for Pragmatic Features

2.1.1 Fine-Tuning and Adaptation of Pre-trained Language Models

Pre-trained language models based on the Transformer architecture form the core technical foundation for pragmatic feature extraction. Through a domain-adaptive fine-tuning strategy, the generic language model is adapted to the specific task of recognizing pragmatic features in learner language. Specifically, a joint training objective combining dynamic masked language modeling and next sentence prediction is employed to enhance the model's understanding of pragmatic contexts. For speech act classification tasks, a task-specific layer is introduced atop the BERT model, utilizing hierarchical transfer learning techniques to capture nuanced differences in pragmatic function realization while retaining general linguistic knowledge. For identifying indirect speech acts, an attention head visualization analysis mechanism is designed to interpret the key evidence for model decisions, ensuring transparency and interpretability in the feature extraction process.

2.1.2 Sequence Labeling and Structural Prediction Models

Formulaic expressions and discourse organizational features in pragmatic competence require sequence-level modeling for effective capture. This study employs a hybrid architecture based on Conditional Random Fields and Bidirectional Long Short-Term Memory networks to achieve boundary identification and functional classification of pragmatic markers, discourse particles, and rhetorical structures. For speech act sequences in dialogues, a graph neural network-based dialogue structure parser is designed, representing conversations as graph structures composed of speech act nodes and pragmatic relationship edges. This approach models pragmatic coherence between turns through message-passing mechanisms, enabling simultaneous capture of local pragmatic features and global dialogue structures, thereby providing multi-level indicators for analyzing learners' discourse organizational competence.

2.1.3 Deep Semantic Representation and Metaphor Computation

The extraction of implicit pragmatic features requires transcending surface-level syntactic analysis to delve into semantic dimensions. Utilizing sentence embedding models based on contrastive learning, this approach maps learner discourse into semantic space and identifies prototypical expression patterns of different pragmatic functions through cluster analysis. For understanding pragmatic presuppositions and conversational implicatures, a commonsense knowledge-enhanced reasoning model is designed, integrating external knowledge bases with contextual information to compute the adaptation degree between discourse and context. Regarding the comprehension of rhetorical phenomena such as metaphor and irony, a multi-task learning framework is adopted to jointly train literal meaning recognition and rhetorical intent classification, thereby enhancing the model's sensitivity to non-literal language use[3].

2.2 Pragmatic Pattern Recognition Techniques in Multimodal Data

2.2.1 Cross-Modal Representation Learning and Feature Alignment

The primary challenge in multimodal pragmatic analysis lies in achieving unified representation of different modal data. This study employs an encoder-decoder architecture based on deep neural networks to process textual, audio, and visual data respectively. For the audio modality, a hybrid feature extractor combining convolutional neural networks and recurrent neural networks is designed to automatically learn prosodic, rhythmic, and timbral features from raw waveforms. Visual modality processing utilizes 3D convolutional neural networks to capture spatiotemporal dynamic characteristics of facial expressions and body movements. To address inter-modal asynchrony, a cross-modal attention alignment mechanism is designed to dynamically learn correspondences between different modal

segments, establishing a joint cross-modal embedding space.

2.2.2 Multimodal Fusion and Collaborative Analysis

Building upon feature alignment, the research compares the effectiveness of various fusion strategies—including early fusion, intermediate fusion, and late fusion—in pragmatic recognition tasks. A tensor fusion network-based multimodal fusion method is innovatively proposed, which explicitly models inter-modal interactions through outer product operations to capture cross-modal synergistic patterns such as "smiles accompanied by soft tones." For specific pragmatic function recognition, a task-adaptive fusion gating mechanism is designed to dynamically adjust the weight of each modality in the final decision-making. For instance, when recognizing sarcastic intent, the model automatically assigns higher weights to prosodic and facial expression features to capture inconsistencies between literal meaning and paralinguistic signals.

2.2.3 Multimodal Pragmatic Graph Construction

To achieve comprehensive assessment of learners' pragmatic competence, the study proposes the concept of a multimodal pragmatic graph. This graph organizes learners' multimodal pragmatic performances across different contexts into structured representations, where nodes represent instances of pragmatic acts and edges denote functional relationships between acts. Through deep learning on pragmatic graphs using graph neural networks, personalized pragmatic style patterns and their developmental trajectories can be identified. This technical framework not only supports static evaluation of pragmatic competence but also reveals the dynamic developmental process of pragmatic ability through temporal graph analysis[4].

2.3 Application Strategies of Machine Learning Algorithms in Pattern Mining

2.3.1 Unsupervised Learning and Exploration of Group Heterogeneity

When dealing with unannotated natural learner corpora, unsupervised learning strategies are employed to explore the natural grouping structure of pragmatic competence. Based on pragmatic feature matrices, Gaussian Mixture Models and spectral clustering algorithms are applied to identify different prototypes of pragmatic competence development within learner populations. To determine the optimal number of clusters, a model selection criterion combining silhouette coefficient, Calinski-Harabasz index, and pragmatics theory guidance is designed. Through visualization techniques such as t-SNE, high-dimensional pragmatic features are projected into a two-dimensional space to intuitively display distribution patterns and developmental gaps within learner groups. Discriminative feature analysis of clustering results can reveal typical pragmatic feature combinations across different developmental groups, providing basis for personalized instructional interventions.

2.3.2 Supervised Learning and Developmental Stage Prediction

For datasets with annotated developmental stages, ensemble learning-based prediction models for developmental levels are constructed. Gradient Boosting Decision Trees and Random Forest algorithms are employed to establish mapping relationships from multi-dimensional pragmatic features to developmental levels. Through model interpretation techniques including permutation importance and SHAP values, the contribution degree of each pragmatic feature to developmental stage discrimination is quantified, thereby identifying key developmental indicators. Addressing the ordinal nature of developmental stage prediction, an ordinal regression framework is designed to constrain model outputs to comply with the progressive characteristics of developmental stages. This prediction model not only enables automated assessment of pragmatic competence but, more importantly, reveals critical pathways and transition points in pragmatic competence development through analysis of discriminative features[5].

2.3.3 Temporal Modeling and Developmental Trajectory Analysis

To capture the dynamic developmental process of pragmatic competence, specialized temporal modeling techniques for longitudinal data are required. Long Short-Term Memory network-based sequence models are adopted to model and predict learners' pragmatic development trajectories. Combined with Hidden Markov Models, latent states in pragmatic competence development and their transition probabilities are identified, revealing phased characteristics of development. To further analyze factors influencing development, a multi-task learning-based causal discovery framework is designed to simultaneously model developmental trajectories across multiple pragmatic dimensions and their interrelationships, exploring potential lead or lag effects between dimensions. These temporal

analysis techniques collectively constitute a computational microscope for understanding the dynamic developmental mechanisms of pragmatic competence.

3. Validation and Instructional Application Research on Pragmatic Competence Development Patterns

3.1 Validity Verification Design for Developmental Patterns

To ensure the scientific validity and reliability of the development patterns mined through machine learning, this study constructs a multidimensional cross-validation framework. The framework first employs internal validity testing, utilizing bootstrap aggregation and cross-validation techniques to assess the stability and generalization capability of the patterns themselves. It calculates confidence intervals for key pattern parameters across different data subsets to determine whether they reflect universal regularities within the learner population rather than accidental characteristics of specific samples. External validity testing involves conducting convergent validity analysis by comparing the mined development patterns with existing pragmatic competence assessment results obtained through traditional measurement methods. Simultaneously, domain experts are invited to evaluate the conceptual congruence of the developmental pathways depicted by the patterns, achieving mutual verification between data-driven discoveries and theoretical experience. Construct validity testing focuses on analyzing the internal structure of the patterns, employing statistical methods such as confirmatory factor analysis to examine whether the predefined dimensions — including linguistic pragmatics, socio-pragmatics, and cognitive processing — demonstrate the anticipated correlations and discriminability in empirical data. This series of rigorous verification procedures collectively constitutes a reliability and validity assurance system for data-driven developmental patterns.

3.2 Analysis of Compatibility Between Instructional Models and Pragmatic Competence Development

Based on the validated developmental patterns, this study further analyzes their compatibility relationships with different teaching philosophies and methods. The core of the analysis lies in mapping the sequential and phased characteristics revealed by the developmental patterns onto the design logic of instructional sequences. For learners demonstrating a "form precedes function" developmental path in the patterns, the teaching emphasis should progressively transition from accuracy in linguistic forms to flexibility in contextual application. For groups lagging in the socio-pragmatic dimension, the instruction needs to emphasize metapragmatic awareness cultivation in communicative tasks and cross-cultural comparative analysis. The research systematically examines the convergence points between different paradigms — such as task-based language teaching and content-based instruction — and specific developmental trajectories. This includes investigating the correspondence between task complexity and development in the cognitive processing dimension, or analyzing the effectiveness of different feedback strategies in addressing pragmatic fossilization at various developmental stages. This compatibility analysis aims to transcend universal teaching principles and form a developmental pattern-oriented precision teaching theory, thereby providing a scholarly foundation for subsequent interventions[6].

3.3 Empirical Evaluation of Instructional Interventions on Pragmatic Competence Enhancement

To examine the actual efficacy of developmental pattern-based instructional interventions, the research designs a rigorous empirical evaluation scheme. Employing a quasi-experimental design, the study selects pre-tested learners as subjects, groups them according to their respective developmental stages or prototypes, and implements targeted instructional interventions. The intervention content is closely designed around specific weak dimensions or key developmental nodes identified by the developmental patterns. For instance, learners in the "intermediate stage" receive substantial high-variability input and awareness-raising tasks to promote the development of their pragmatic inference ability. The evaluation process comprehensively utilizes quantitative indicators automatically extracted by machine learning models and qualitative discourse analysis conducted by trained raters to conduct multidimensional measurement of changes in learners' target pragmatic features before and after the intervention. After controlling for pre-test differences through statistical models such as analysis of covariance, the net effect of the intervention is precisely estimated. Furthermore, by conducting new rounds of pattern mining on post-intervention learner corpora, the study tracks whether

developmental pathways progress toward higher stages, thereby dynamically evaluating the reshaping effect of instructional interventions on pragmatic competence developmental trajectories. This process ultimately forms an evidence-based research cycle of "mining-validation-intervention-reevaluation."

Conclusion

This study has successfully identified developmental patterns of pragmatic competence in English learners by systematically constructing a theoretical framework for pragmatic competence development, innovatively applying machine learning technologies to achieve automated extraction of pragmatic features and multimodal pattern recognition, and conducting rigorous validity verification and teaching empirical research. The research confirms that machine learning methods can effectively process large-scale complex corpora and reveal developmental regularities and group heterogeneity that are difficult to detect through traditional approaches, thereby providing a new methodological paradigm for research on pragmatic competence development. The established theoretical framework and technical pathway not only deepen the understanding of the multidimensional dynamic development process of pragmatic competence but also provide practical guidance for developing data-driven precision teaching intervention programs. Future research could further advance in the following aspects: expanding the scope and quality of multimodal data collection while incorporating richer contextual factors and socio-cultural variables; developing more explanatory temporal models to capture nonlinear characteristics of development; exploring universal patterns and specific manifestations of pragmatic competence development across different linguistic backgrounds; and promoting the transformation of research findings into adaptive learning systems and intelligent tutoring tools, ultimately constructing a research ecosystem for pragmatic competence development where theoretical innovation and technological application mutually reinforce each other.

References

[1] Ma, Ying. "An Investigation into the Developmental Status of Chinese EFL Learners' Pragmatic Competence Based on the China Standards of English." Proceedings of the First International Forum on Frontier Linguistics, Literature, and Teaching & Young Teachers' Educational Development Forum (II). Ed. Shenyang Ligong University;, 2023, 50-57.

[2] Kang, Yujing, and Wu Qingran. "Investigation and Analysis of Interlanguage Pragmatic Competence among College English Learners." Modern English 12(2023):57-60.

[3] Zhang, Xiao. The Impact of Face-to-Face and Synchronous Computer-Mediated Communication on the Acquisition of Refusal Strategies by English Learners. 2023. Central South University, MA thesis.

[4] Lin, Haiming. Research on Influencing Factors of Key Competencies in College English Learners. 2022. Southwest University, PhD dissertation.

[5] Xiao, Huan, et al. "Investigation and Research on Pragmatic Competence of Foreign Language Majors—Based on the China Standards of English." Campus English 47(2021):29-30.

[6] Wang, Yuqi. An Empirical Study on Formulaic Language Pragmatic Competence of Chinese EFL Learners. 2021. East China Normal University, PhD dissertation.