International Journal for Housing Science and Its Applications
Publish August 10, 2025. Volume 46, Issue 4 Pages 6628-6639

https://doi.org/10.70517/ijhsa464563

The Linguistic Presentation of the Symbolism of Residential Space in English Literature

Yu Zhao[1] and Yunxia Yan[2,*]
[1] Foreign Language Teaching Department, Changzhi Medical College, Changzhi, Shanxi, 046000, China
[2] Translation Department, Hebei University of Science & Technology, Shijiazhuang, Hebei, 050018, China
Corresponding authors: (e-mail: zhao2yu200@126.com).

Abstract As an important carrier of cultural metaphor and human portrayal in English literature, the symbolism of residential space is often presented through complex linguistic forms and narrative structures. In this study, we propose an innovative approach that integrates literary criticism theory and natural language processing technology, and construct the Attention-BERT sentiment analysis framework based on the attention mechanism and the BERT model, aiming to systematically analyze the multidimensional symbolism of houses in English literature. Three core analysis clues are refined at the theoretical level. On the technical level, the Attention mechanism is introduced to dynamically weight the key semantic units, and the BERT bidirectional coding capability is utilized to parse the masked language, realizing the end-to-end mapping from text to sentiment tendency. The experiments validate the model performance by comparing the baseline models such as LSTM, RoBERTa, etc. Attention-BERT takes a significant lead with an accuracy of 83.53% and an F1 value of 80.05%, which is an improvement of 10.49 percentage points compared to the 73.04% of the traditional LSTM, and verifies the validity of combining the attention mechanism with the pre-trained model. Further quantitative analysis shows that the residential spatial text presents high diversity in lexical features, with a class character shape character ratio of 12.01% versus a high density feature, and a 67.30% share of real words. The frequent use of nouns (29.15%), verbs (20.91%) and adjectives (9.50%) strengthened the spatial figurative and metaphorical expression. At the syntactic level, the significant high-frequency distributions of quadratic lexical clusters (NSPC, VSPC), with all LLR values >21.94, $p<0.01$, reveal the thematic association of space with power and identity.

Index Terms residential space, English literature, attention mechanism, BERT model, sentiment analysis

I. Introduction

Dwellings provide spaces for human beings to inhabit and to carry out the practice of social life; they shelter, preserve, and participate in the creation of perceptions, memories, and imaginaries belonging to their occupants, bridging architecture, occupant's bodies, and cultures, and providing multiple possibilities for thematic, narrative, and stylistic writing about space in English-language literature [1]-[4]. The concept of “home” encompasses both “external” architectural forms that house individuals or families, as well as “internal” spatial forms that showcase the experience of daily life and intimate experiences. In the practice of individuals and groups in society, residential experience is a common echo of family relations, social structure and architectural space, an intermediate level of individual life feeling and social culture [5], [6]. Especially in the context of modern society, the residential space and its experience in English literature can become a special metaphor that suggests the deeper meaning of the textual existence in connection with the narrative mode, cultural context and group mind [7], [8].

On the one hand, the modern house contains an individual space full of hidden experiences, and in correspondence with the public space it becomes a place of refuge for frustrated idealists, a space full of muttered monologues. On the other hand, the modern society in transition demands a break with the old cultural order, and so the residential space is able to become a microcosmic realm in which cultural change takes place and is shaped as a space full of dialogicity as a place where ideological and social power struggles take place [9]-[11]. In the presence of such contradictions, the portrayal of residential space and its experiences in English-language literature forms cultural forms that contain cultural metaphors and symbols [12]. How, then, does the field of contemporary literature view the linguistic presentation of the symbolism of residential space in literary works.

This paper proposes an innovative method integrating literary criticism theory and natural language processing technology to reveal the multidimensional symbolic meaning of residential space in the text by constructing an affective analysis model integrating the attention mechanism. Firstly, based on modern literary criticism theory, three core analytical clues are sorted out to provide theoretical support for the literary interpretation of residential space.

Specifically, through the perspective of “body-dwelling-power”, we analyze the mapping function of residential space on social structure; through the correlation of “symbol-perception-space form”, we explore the interactive mechanism between spatial symbols and group identity; and through the combination of “individual-space archetype” and “individual-space archetype”, we analyze the multidimensional symbolic significance of residential space in the text. With the connection of “symbol-perception-space form”, the article explores the interaction mechanism between spatial symbols and group identity; and combines the “individual-space archetype-group mind” to explore the archetypal significance of residential space in the cultural transformation. On this basis, the article introduces deep learning technology to realize the vectorized characterization and sentiment analysis of the text. Through the attention mechanism, the model can focus on the key semantic units related to residential space in the text, and dynamically capture the contextual associations of various literary imagery by calculating the similarity weights of query vectors and key vectors. Further, in combination with the pre-trained BERT model, its bidirectional encoding capability is utilized for deep parsing of the masked language, and the Attention-BERT fusion model is finally constructed to realize the end-to-end mapping from literary texts to emotional tendencies.

II. Research on the method of literary spatial sentiment analysis based on the attention mechanism and the BERT model

II. A. Three Literary Critical Thoughts on Residential Space in English Literature

The depiction of residential space and its experience in modern English literature is very rich, and the perceptual experience related to modernity revealed therein has formed a unique cultural value in terms of narrative, psychology, ideology, etc., and has become a valuable object of literary criticism. The depiction of residential space and its experience in modern literature is not a “depiction of residence” or “space writing” in the general sense, but an integral part of the literary text, carrying the cultural contradictions originated from modernity.

The depiction of residential space and its experience in modern literature forms a cultural form that contains cultural metaphors and symbols, and in terms of context and methodology, it is related to the problematic consciousness of researchers who hold different critical ideas, specifically, the following three critical ideas.

II. A. 1) The “body-dwelling-power” trail

Residential space is a mapped physical form of the human sense of self-presence, and is also associated with or a metaphor for human socio-cultural relations and power mechanisms. Residential space and its experience is a projection of social structure and social relations, and an intermediary for the interaction and transformation of human residential behavior with macro-social structure. Residential space and its experiences are objects to be consumed in literary texts, but they are also rich in narrative potential. On a more fundamental material level, the architectural entity on which residential space is based is also a special means of production and an object of consumption, and its production and consumption are also the subject of modern literary works. Literary critics can regard the residential space in modern literature as a product of social power, which can be used to examine the role of power mechanisms in the shaping of modern social space, and lead to reflections on modern Western culture, the form of residential space, and the types of experience.

II. A. 2) The “symbol-perception-space-form” trail

A series of systematic symbols and meaningful spatial forms constructed by residential space in literary works, which correspond to the existential situation, cultural interests and identities of the occupants, can also be regarded as the clue of “symbol - spatial form - identity” by literary critics. This can also be emphasized by literary critics as the clue of “symbol - spatial form - identity”. In terms of the presentation of social relations, the residential space in the modern city is not an existential object free from emotions, but its form continues the logic of social economy and the order of social class, so that differences and contradictions are formed in different residential spaces. In the history of literary and cultural criticism, the modern audience's experience of literary works has been linked to the perception of “urban texts” such as Paris, which have been presented, described and analyzed in a “literary montage”, leading to the proposal of four typical urban spaces The four typical scenes of urban space are presented: the arcade street, the Western scene, the World's Fair, and the personal apartment. The “personal apartment”, as a form of private residential experience, is a cultural form between the individual and the city, and has a special value for literary criticism. This kind of thinking is a profound inspiration for the criticism of the form of residential space and its experience in English literature.

II. A. 3) “Individual - Spatial Archetype - Group Mind” Clues

The home carries the intimate feelings that originate from the body, which is in common with the narrative means of literature dedicated to displaying the cultural experience of human groups. The residential space, which communicates between the individual and the group mind, nurtures a matching perceptual structure that participates

in social and cultural changes. The interior of modern residential space is not homogeneous; it is often designed and divided into zones with different functions, and the experience of the occupants varies according to the function of the space. On a more macroscopic level, residential space and its experience in literature of the age of cultural transformation are inextricably linked to many cultural forms of coloniality and modernity in modern society, forming spatial archetypes that are deeply embedded in literary imagery. In addition to the clues between the dwelling and the social structure in literature, the relationship between the dwelling and the portrayal of human nature can also attract the attention of literary critics.

II. B.Text Vectorization Techniques - Attention Mechanisms

The above three literary criticism clues reveal the multidimensional symbolic meanings of residential space in the text, however, how to systematically extract these implicit associations from the massive literary corpus still requires the use of technological means to realize the refinement of the text. To this end, this chapter introduces the attention mechanism, which focuses on the key semantic units through a dynamic weighting strategy to provide technical support for the quantitative analysis of literary symbols.

In deep learning tasks, the attention mechanism can make the model focus more on the important parts when processing data, thus improving the performance of the model. The core idea is that when the model processes the input data, it does not simply treat all information equally, but gives different degrees of attention according to the importance of each input, and the structure of the attention mechanism is shown in Figure 1.

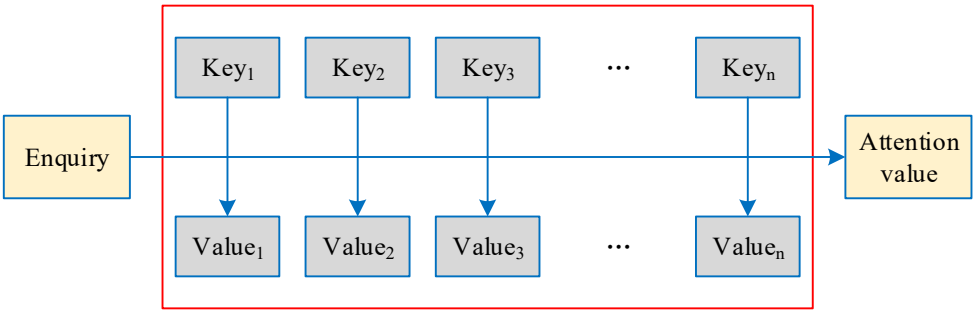

Figure 1: Attention mechanism

The Source of the attention mechanism is a collection of Key-Value data pairs, and the Value in the figure represents the similarity weight of the Query with the corresponding Key. The final attention value is obtained by weighting and summing the similarity between each Query and Key, which is calculated as shown in equation (1).

$$Attention(Q,S)=\sum_{i=1}^{l_s} similarity(Q,K_i)\times V_i \tag{1}$$

where Q denotes the query vector, K denotes the key vector, V denotes the value vector, S denotes the Source, and $l_x=\|S\|$ is used to measure its size in the space.

The computational process of the attention mechanism consists of three steps. First, the similarity between Q and K is calculated using similarity metric functions, commonly used functions include dot product, scaled dot product, weighted dot product, etc. The similarity score of query and key is calculated as shown in equation (2).

$$Similarity(Q,K)=Q\cdot K \tag{2}$$

Subsequently, based on the computed similarity scores, the similarity scores are converted into attention weights by applying a Softmax function. Where the Softmax function ensures that the attentional weights sum to 1, indicating the weight corresponding to each key. Assuming that the similarity score is S, the attention weight matrix A is calculated as shown in equation (3).

$$A=Soft\max(S) \tag{3}$$

Finally, a weighted summation is performed using the attention weights to obtain the final attention value.

With the above steps, the model is able to weight the values based on the similarity between the query and the key to obtain a more informative output. The calculation of the attention output is shown in equation (4). This formula can be understood as first calculating the similarity between the query vector Q and each key vector K^T

(calculated by inner product), then converting these similarities into weights by Softmax function, and finally using these weights to weight and sum the value vectors to obtain the final attention output. This can effectively solve the problem of vanishing or exploding gradients, making the training of the model more stable and effective.

$$Attention(Q,K,V) = Soft\max\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{4}$$

where d_k denotes the dimension of the vector.

II. C.BERT Sentiment Analysis Model

Although the attention mechanism can effectively capture local semantic weights, the analysis of deeper sentiment tendencies of literary texts still needs to be combined with more powerful semantic representation models. Based on this, this chapter further adopts the BERT model to construct a classification framework for residential spatial sentiment analysis through pre-training and fine-tuning strategies, realizing the full-process mapping from symbol parsing to sentiment prediction.

Text sentiment analysis is essentially a task of revealing textual tendencies through classification, in which the supervised learning approach of machine learning is proven to be effective in performing this task. While this approach is effective, it relies on extensive manual annotation work, requiring domain experts to first perform careful sentiment labeling of the text to construct a training sample set, and then apply traditional machine learning models such as Support Vector Machines (SVMs), Random Forests, or the more advanced BERT model for sentiment analysis. In this paper, the BERT model is chosen to predict affective tendencies.

The special feature of the BERT model is its bi-directional encoder structure, which realizes the possibility of deeper understanding of the text by improving the Transformer and introducing two pre-training tasks, namely, Masked Language Model (MLM) and Next Sentence Prediction (NSP).

In the first step, the inputs to the BERT model are determined. Specifically, the input processing of BERT is very unique. It not only processes sentences after word separation, but also inserts special tokens (e.g., [SEP] separator and [CLS] categorical tokens) between sentences as a way of training the model, so that the [CLS] tokens can ultimately imply the sentiment features of the overall text, which can then be used for sentiment classification. The input to the model consists of three parts: word vectors, segmentation embeddings, and positional coding, which work together in the Transformer structure to achieve a deeper understanding of the text and sentiment prediction, and the input to the BERT model is shown in Figure 2 below.

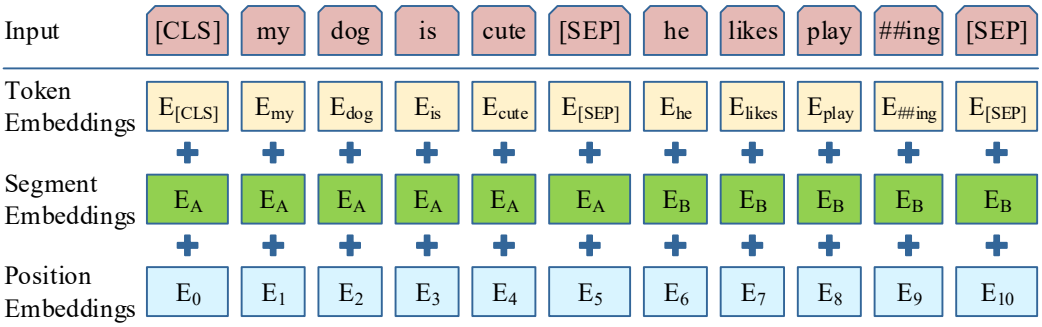

Figure 2: BERT model input

For computational convenience, all three vector dimensions are e in BERT, so the input representation v corresponding to the input sequence can be computed by the following equation:

$$v = v^t + v^s + v^p \tag{5}$$

where v^r denotes the word vector; v^s denotes the block vector; v^p denotes the position vector; the size of all three vectors is $N \times e$, N denotes the maximum length of the sequence, and e denotes the dimensionality of the word vector.

In the second step, the BERT model is pre-trained. In the masked language model (Mask-LM) training phase of the BERT model, a unique strategy is to randomly select about 15% of the words in the text to be masked. Specifically, 80% of these words are replaced with special $[mask]$ tokens, 10% are randomly replaced with other words in the text, and the remaining 10% are left as they are. The goal of the method is to use context to predict

the words that are masked (or replaced, or left as is), and iteratively optimize the model parameters by back-propagating the prediction loss for the $[mask]$ position. Given that the BERT model itself is extremely large, the computational resources and time costs required to train a BERT model directly from scratch are enormous. Therefore, this paper adopts a migration learning strategy by choosing a pre-trained model - i.e., bert-base-chinese - as a starting point. This pre-trained model has been trained on an extensive Chinese corpus, thus learning rich linguistic features and patterns.

In the third step, a model for categorizing the sentiment of comments in English literature is trained. The model was fine-tuned by feeding the collected texts about residential spaces into the model as a way to adapt it to specific sentiment analysis tasks. Ultimately, the performance improvement and optimization effect of the model is evaluated by testing these fine-tuned model parameters on an independent validation set, and the flow of the text fine-tuning steps is shown in Figure 3.

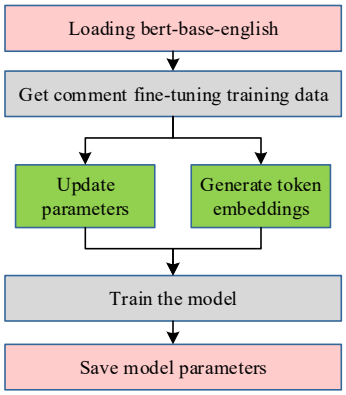

Figure 3: Text fine-tuning steps

After the optimizer and loss function of the model are set explicitly, iterative training of the model can be started. In this paper, cross entropy is used as the loss function for model training, and its formula is expressed as follows:

$$loss = -y\log(\hat{y}) - (1-y)\log(1-\hat{y}) \tag{6}$$

In the fine-tuning and prediction part, the input text is processed through the Tokenizer, where the text is first separated into individual characters up to a specified length, then mapped to the corresponding ID, and the first bit is added with a specific sign character to get the final Token Embedding.The Token Embedding is combined with the Segment Embeddings and the The Token Embedding, together with the Segment Embeddings and PositionEmbeddings, form the incoming data to the model. The input data is first encoded with features through a multilayer bi-directional transformer, and then the obtained features are pooled and modified to the specified shapes and passed to the fully connected layer for classification, the BERT classification process is shown in Fig. 4.

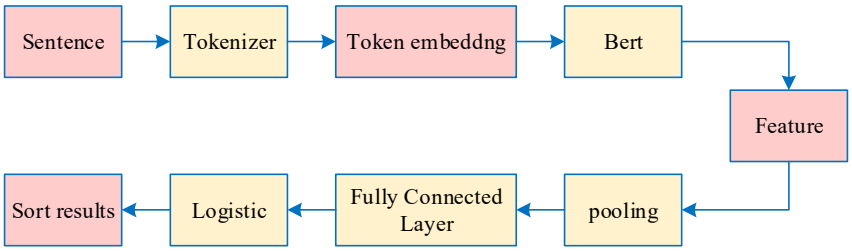

Figure 4: BERT classification process

An Attention-BERT sentiment analysis model incorporating the attention mechanism is finally synthesized.

III. Comparative model analysis and quantitative analysis of the symbolism of residential spaces

The Attention-BERT model constructed in Chapter 2 provides a technical framework for the sentiment analysis of residential space, but the deeper analysis of its cultural metaphors needs to be further combined with linguistic formal features. Therefore, Chapter 3 focuses on the quantitative analysis at the lexical and syntactic levels, and reveals the systematic preference of residential space texts in terms of language selection by comparing the empirical data from the observation library and the reference library, so as to underpin the mechanism of generating their symbolic meanings.

III. A. Emotional Analysis of Residential Elements in English Literature

III. A. 1) Experimental environment

A computer equipped with an AMD Ryzen 9 5900HX CPU, 32 GB of RAM, and an NVIDIA GeForce RTX 3080 Ti GPU was used in this experiment for model training and evaluation.

Considering the chosen model structure, this paper performs parameter tuning for the combination of the knowledge graph and the BERT model that introduces the attention mechanism. The hyperparameters are set as follows: the learning rate is set to 1e-4, the AdamW optimizer is used for tuning the parameters, and the training batch size is set to 32. Since the maximum length of the input text is not more than 200 English words, the maximum length of the input sequence of the BERT is set to 200. The dropout ratio in the prediction layer is set to 0.2 to enhance the robustness of the model. Meanwhile, in order to prevent overfitting, this paper also introduces L2 regularization with a regularization coefficient of 1e-5. In the experiments, 50 epochs are preset as the maximum number of rounds of training, and in order to avoid overfitting phenomenon in training, this paper introduces the method of early stopping. When the loss function value of the model on the validation set continues to increase for five consecutive times, the training of the model is stopped, and the model with the smallest loss function value on the validation set is saved for testing.

III. A. 2) Data sets

In order to comprehensively assess and validate the applicability and effectiveness of the proposed sentiment analysis model in the English context. The dataset used in this study consists of two parts: first, text fragments containing descriptions of residential space manually extracted from English literary classics, covering 50 representative novels and plays from the 19th to the 21st centuries; and second, readers' comments crawled from platforms such as Goodreads, Amazon Books, and so on, focusing on discussions of the symbolic significance of residential space in the literary works. The dataset contains a total of 12,000 text samples, each of which is controlled to be between 50-200 words in length to ensure the normality of the model inputs.

The dataset is divided into a training set (8400 entries), a validation set (2400 entries), and a test set (1200 entries) in the ratio of 7:2:1, and is divided in such a way as to ensure a balanced distribution of works from different periods and genres.

III. A. 3) Baseline model

In order to comprehensively assess the performance of Attention-BERT, the model proposed in this paper, and to ensure the reliability and objectivity of the research results, this study conducts a comprehensive comparative analysis of the current popular deep learning-based text sentiment classification methods.

Given that this paper focuses on implicit sentiment analysis methods, implicit sentiment analysis models that incorporate external knowledge in recent years are specifically chosen as comparison benchmarks. The specific comparison methods are as follows:

LSTM: Adopting long and short-term memory network to capture temporal information in text, it alleviates the problem of gradient vanishing and is suitable for dealing with long-distance-dependent scenes in text.

Attention-LSTM: Combines the attention mechanism on the basis of LSTM, assigns differentiated weights to different parts of the text, and enhances the model's ability to recognize key emotional information in the text.

BERT: As a pre-training model based on bidirectional encoder representation, it utilizes large-scale corpus pre-training to deeply understand the text semantics, which is especially suitable for complex sentiment analysis tasks.

RoBERTa: as an optimized version of BERT, it improves the training efficiency and performance by improving the pre-training process, which is especially suitable for detailed sentiment analysis.

Context-BERT: not only combines the advantages of BERT's bi-directional encoder, but also effectively recognizes sentences that do not contain obvious emotion words but still convey emotion through the contextual query focus mechanism.

SCL-BERT: adopts a supervised contrast learning method, which is based on sentiment labels during the training process, and strengthens the average embedding distance between texts with different sentiment labels, thus improving BERT's ability to recognize implicit sentiment.

III. A. 4) Loss training

In this paper, the Attention-BERT model was first trained in the dataset for in-depth learning, and the loss function values for each iteration cycle were recorded, and the training loss curve is shown in Fig. 5.

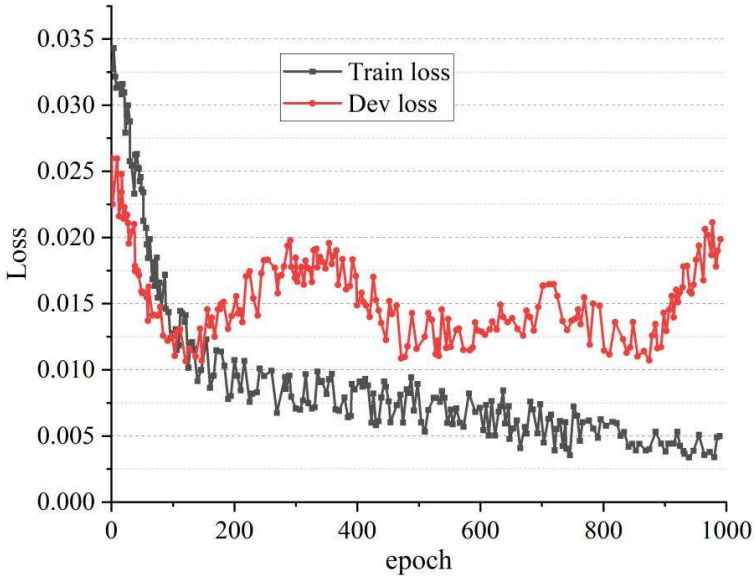

Figure 5: Training loss curve

By analyzing the loss curves of the dataset in depth, a clear trend can be observed: as the number of training iterations increases, the loss of the training set gradually decreases and gradually stabilizes. However, at the same time, the loss of the validation set shows a gradual increase, indicating the occurrence of overfitting phenomenon. In order to avoid the negative impact of overfitting on the generalization ability of the model, the model with the lowest loss on the validation set is preserved in this paper during the training process.

III. A. 5) Comparative experimental analysis of different models

The aspect sentiment analysis studied in this paper is essentially a categorization task, and considering the characteristics of the categorization task, in the experiments this paper chooses the accuracy rate and F1 value as the evaluation index. Each model conducts 10 experiments on the dataset, and the average comparison results of this paper's model and six comparative baseline models are shown in Fig. 6.

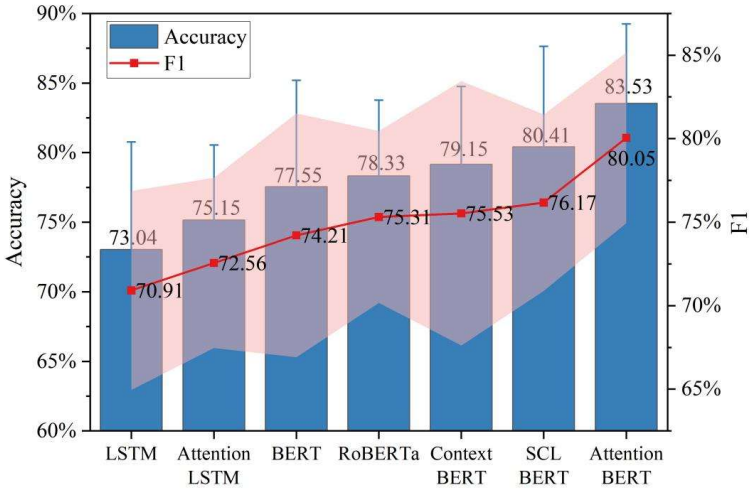

Figure 6: Comparison of the accuracy and F1 value

The experimental results show that Attention-BERT significantly outperforms other models with an accuracy of 83.53% and an F1 value of 80.05%, which is an improvement of 10.49 percentage points compared to the traditional LSTM model (accuracy of 73.04%), which verifies the effectiveness of combining the attention mechanism with the pre-trained model. Specifically, Attention-LSTM has an accuracy of 75.15%, which is a 2.11% improvement over the base LSTM, indicating that the attention mechanism can enhance the model's ability to capture implicit emotional cues by weighting the key semantic units.The higher performance of BERT's accuracy of 77.55% and its optimized version RoBERTa's of 78.33% reflect the bi-directional coding structure's ability to parsing of complex literary metaphors.

Attention-BERT further integrates local focus and global semantic representation by embedding the attention layer into the Transformer architecture of BERT, and its F1 value is improved by 3.88% compared with the 76.17% of SCL-BERT, which indicates that the dynamic weight allocation can effectively alleviate the problem of obscurity of emotional expression in literary texts.

III. B. Quantitative analysis of the symbolic meaning of residential space: a study based on lexical and syntactic features

After completing the performance validation of the sentiment analysis model, in order to further understand how the linguistic forms of residential space texts serve their symbolic functions, this section starts from the lexical features, and analyzes the semantic density and metaphoric potential of the texts through the quantitative comparison of the class of character-form character ratios and lexical density.

III. B. 1) Data processing

The experimental data in this section are divided into an observation bank and a reference bank. The observation pool consists of literary works embodying residential space. 5,000 fragments containing typical descriptions of residential space were selected from the 12,000 text samples described in Section 3.1, covering English literature from the 19th to the 21st centuries, and encompassing different time periods, genres, and authorial styles.

The reference library includes 5,000 English literary works of various genres except descriptions of residential spaces.

Data preprocessing: Text cleaning: remove punctuation, numbers and non-literary annotations (e.g., editor's notes, footnotes);

Segmentation and lexical annotation: Segmentation using the NLTK tool and lexical annotation via Stanford CoreNLP to distinguish between real words (nouns, verbs, adjectives, adverbs) and imaginary words;

Deactivation filtering: eliminating high-frequency function words (e.g., articles, prepositions) to focus on key semantic words.

III. B. 2) A Study of the Lexical Characteristics of Residential Space in English Literature

Lexical richness is the primary indicator for examining the macro-lexical characteristics of literary texts, which can be categorized into two main examination panels, namely, lexical diversity and lexical density. The above observational variables can be reflected by calculating the ratio of class symbols to form symbols, the proportion of real words, and the use of disposable words, respectively.

(1) Lexical Diversity

The number of shape symbols is the total number of all words in the corpus, while the number of class symbols counts repeated words only once. The class symbols to shape symbols ratio mainly reflects the degree of lexical diversity of the text, i.e., the higher the ratio is, the lower the repetition rate of the words, the stronger the lexical diversity and richness of the text, and the more difficult it is to read the text.

Table 1 lists the lexical feature indicators of the two corpora.

Table 1: Lexical feature indicators of the two corpora

	Tokens	Types	TTR	STTR
Observation database	228342	27422	12.01%	58.61%
Reference database	1283306	93234	7.27%	46.24%

Table 1 compares the lexical diversity metrics of the observation library (residential spatial texts) with the reference library (non-residential spatial texts). The results show that the TTR of class-symbol-to-form-symbol ratio of the observation library is 12.01%, which is significantly higher than that of the reference library, which is 7.27% ($p<0.001$), suggesting that the lexical repetition rate of residential spatial texts is lower and their diversity is higher. For example, the number of class symbols in the observation library is 27,422 (total form symbols 228,342), while

the number of class symbols in the reference library is 93,234 (total form symbols 1,283,306), and the TTR of the reference library is only 60% of that of the observation library, despite its larger size. This difference highlights the tendency of residential spatial texts to enhance symbolic expression by enriching lexical choices.

Further analysis of the STTR of the standard class symbol-to-shape ratio shows that the STTR of the observation library is 58.61%, much higher than that of the reference library, which is 46.24%.The significant difference in the STTR, which eliminates the interference of the text length on the TTR through segmentation standardization, further validates the lexical complexity of the residential spatial texts.

(2) Lexical Intensity

In order to further validate the relevant conclusions obtained based on data such as the standard class of character shape and character ratio, lexical density is the percentage of the number of real words in a text out of the total number of words, the lexical density of the study text was examined to further expand the significant features of English literature embodying the residential space at the level of lexical density. The frequency of use of each word class in the text can be obtained more easily by utilizing the text that has been lexically encoded, and the following comparison of lexical density between the observation and reference pools can be obtained Table 2.

Vocabulary in English can be categorized into ten major word classes according to lexical properties, which are nouns, verbs, pronouns, prepositions, adjectives, adverbs, numerals, articles, conjunctions and exclamations.

Table 2: Comparison of vocabulary density between two databases

Word class	Observation database		Reference database		t	p
	Frequency	Percentage	Frequency	Percentage		
Verb	41302	20.91%	192696	18.06%	4.21	<0.001
Noun	57586	29.15%	256661	24.06%	6.34	<0.001
Adjective	18767	9.50%	64167	6.01%	5.12	<0.001
Adverb	15267	7.73%	51332	4.81%	4.78	<0.001
Preposition	22734	11.51%	179663	16.84%	-3.89	<0.001
Pronoun	11317	5.73%	102764	9.63%	-5.67	<0.001
Numerals	2283	1.16%	12833	1.20%	0.15	0.881
Article	18667	9.45%	141264	13.24%	-4.56	<0.001
Conjunction	9138	4.63%	64165	6.01%	-2.45	0.014
Interjection	456	0.23%	1283	0.12%	1.22	0.22
Number of content words	132922		564856		-	
Total number of words	197517		1066828		-	
Vocabulary density	67.30%		52.95%		9.87	<0.001

From Table 2, it can be seen that the total proportion of real words (nouns, verbs, adjectives, adverbs) in the observation database of residential space text is 67.30%, which is significantly higher than that of the reference database (52.95%) (t=9.87, p). <0.001). This difference suggests that the description of residential space relies more on concrete, concrete vocabulary to characterize spatial characteristics. For example, the frequent use of nouns (29.15% vs. 24.06%) may correspond to spatial entities such as "attic", "corridor", and "basement"; The richness of adjectives (9.50% vs. 6.01%) implies a detailed depiction of spatial attributes, such as "wet", "oppressive", and "bright".

The percentage of fictitious words (prepositions, articles, pronouns, etc.) in the observed pool is 32.70%, which is lower than the 47.05% in the reference pool. The reduction of prepositions (11.51% vs. 16.84%) and articles (9.45% vs. 13.24%) is particularly significant (p<0.001), suggesting that residential spatial texts are more inclined to construct scenes directly through real words rather than relying on grammatical function words. This is consistent with the need for literary symbolism to be conveyed through concrete imagery.

The significant increase in the proportion of verbs (20.91% vs. 18.06%) in the text of residential space (t=4.21, p<0.001) may reflect the interaction between the characters' behaviors and the residential space, such as "pushing open the door" and "curling up in the corner", suggesting that the space restricts and shapes the character's psychology or behavior.

The lexical density of the observed pool, 67.30%, is much higher than that of the reference pool, 52.95%, indicating that the semantic information of the residential space texts is denser. This high-density characteristic is closely related to the multi-level expression of literary symbolism - through the dense combination of real words, the text is able to convey both spatial features and metaphorical connotations.

Residential spatial texts show a significant preference for real words and a high lexical density in their vocabulary choice, which not only strengthens the figurative nature of spatial depictions, but also provides a linguistic basis for the generation of symbolic meanings. The rich use of verbs and nouns reveals the dynamic connection between space and behavior, while the high frequency of adjectives deepens the expression of emotion and metaphor. These lexical features together support the multiple symbolic functions of residential space in English literature.

III. B. 3) A Study of the Syntactic Characteristics of the Residential Space in English Literature

Lexical features reveal the semantic denseness of residential spatial texts, but the transmission of symbolic meanings needs to rely on the support of syntactic structures. This section explores how syntactic patterns construct thematic associations of space with power and identity by extracting quaternary lexical clusters with log-likelihood analysis, thus presenting a complete picture of the multi-level interaction between linguistic forms and cultural metaphors.

The goal of syntactic analysis is to analyze input sentences and get their syntactic structure, which is one of the classical tasks in the field of natural language processing. The extraction and analysis of syntactic features mainly focuses on sorting out the syntactic relations of the sentences in the analyzed text language. This paper mainly focuses on the structural analysis of keyword class clusters.

Word class clusters can also be called word blocks, lexical phrases, programmed sequences, phrase structures. As lexical phrases can assume certain sentence components in a sentence, the extraction of high-frequency word class clusters can obtain stable and frequently used multi-word combinations with certain grammatical significance. The collocations of multi-word sequences are mainly based on free combinations of words, special collocations, and fixed idioms and slang. In this chapter, we mainly extract the keyword clusters of the English literary works about residential space, and at the same time, we organize the relevant word cluster structure in the reference corpus of other literary works, and calculate and compare the log-likelihood value in the concept of statistics, so as to analyze the syntactic features of the corpus text from the perspective of lexical co-occurrence relationship.

The word class clusters/N meta clusters of the target corpus can be extracted conveniently through the corpus software Antconc3.5.8. The word class cluster list can clearly and completely show the co-occurrence relationship between words and lexemes, and lead to the significant class connections in the study text, which tends to show particular importance for language research and language application value through the syntactic-semantic categorization of the generated word cluster list. The study of word clusters (chunks) requires the researcher to set the text distribution thresholds, frequency thresholds, and the length of the target chunks on his own, which is somewhat subjective and arbitrary. Examining the distribution of word blocks of different lengths from three words to six words, we can find that the content structure of four-word word blocks is more representative and valuable for research, so this paper mainly includes four-word word blocks in the scope of consideration.

The main operation steps are as follows: import the observation database (cleaned text) into the corpus research software Antconc3.5.8 in UTF-8 encoding format, set the length of the word strings manually, and retrieve the multi-word words or phrases in the corpus. Here, the frequency threshold was set to more than 10 occurrences (the frequency threshold was set to ensure the significance of the extracted multi-word sequences), and the text distribution threshold was set to 3 occurrences (the limited text distribution threshold was set to avoid that multi-word sequences occurring with high frequency in a single text were counted as significant chunks of words). The target word chunks were automatically extracted from the observation corpus by checking N-Grams and clicking Start to run the software to get the 4-element word class cluster table of the observation corpus. The extracted 4-element word cluster list of the research corpus is exported and put into the corpus encoding software Treetagger to encode its lexical properties to get the encoded text (here, it can also be manually identified and organized according to the size of the corpus). The key 4-element word class clusters are shown in Table 3.

The multivariate word clusters generated by the N-gram model can be extracted based on frequency counts and directly observe the stable emergence of multi-word construction information in the text, which can reflect certain local linguistic patterns, present linguistic co-occurrence information, and reflect the probabilistic results of repeated co-option of vocabulary by language speakers (writers) with the support of sufficient corpus. (The lexical assignment marking categories here have been expressed in the previous paper.) According to the structural features of the 4-element lexical clusters (consecutive four-word phrases) in the table, the following categories can be mainly classified: subordinate clauses fragments SPNC, SPNS indicating chronological order; subordinate clauses fragment VVRC indicating reason; subordinate clauses fragments CSPP, CQPQ indicating succession; containing verb fragments VSPC, VVRC, VNCQ, PRQV (with overlap with the other classifications); containing grammatical negatives PRQV, VNCQ, QSPQ; and fragments indicating fixed collocations NSPC, SPSN.

Table 3: Key 4-element word class cluster

4-element word cluster	Frequency of Observation database	Frequency of Reference database	LLR
SPNC	142	63	28.74***
NSPC	98	25	36.12***
VSPC	115	48	24.56***
QSPQ	76	15	42.89***
SPNS	89	32	22.31***
SNSP	54	12	29.67***
CQPQ	67	18	34.15***
CSPP	102	40	26.83***
VVRC	123	55	21.94***
VNCQ	81	22	38.02***
SPSN	45	9	31.45***
PRQV	37	6	40.18***

LLR: Log likelihood ratio values greater than 3.84 indicate a significant difference ($p<0.05$) and *** indicates $p<0.01$.

According to the analysis in Table 3, spatial orientation with noun-dominated structures, NSPC class clusters were significantly more frequent in the observation pool than in the reference pool (98 vs. 25, LLR=36.12***), indicating that residential spatial texts reinforce the physical boundaries and metaphorical meanings of the space through specific orientation vocabulary. Verb-guided spatial behavior also exists, with VSPC-like clusters having a frequency of 115 in the observed library and 48 in the reference library, LLR=24.56***. This type of structure reveals the interaction between the characters and the residential space through the combination of dynamic verbs and spatial entities. For questioning stanzas and spatial metaphors, the QSPQ-like clusters had a frequency of 76 times in the observation pool, 15 times in the reference pool, and LLR = 42.89***. The questioning syntax was frequently used to trigger reflection on spatial symbolism. Temporal order and spatial narrative, SPNC-like clusters were 142 times frequent in the observation pool and 63 times in the reference pool, LLR = 28.74***. The superposition of temporal pronominal and spatial descriptions reinforces the temporal dimension of space, suggesting the influence of historical precipitation or accumulation of events on spatial meaning. Negation structure and spatial contradiction: the frequency of PRQV-like clusters was 37 times in the observation pool and 6 times in the reference pool, LLR=40.18***. The pairing of negations with spatial descriptions highlights spatial closure and oppression, mapping power suppression or identity anxiety.

Overall, these syntactic features collectively serve the symbolic function of residential space, constructing spatial mappings of power relations through complex sentences or channeling critiques of colonial modernity through interrogative sentences. Quantitative analyses show that residential space texts have significant systematic preferences in syntactic choices that provide structural support for their cultural metaphors.

IV. Conclusion

This study systematically reveals the multidimensional symbolism of residential space in English literature and its linguistic generation mechanism by integrating literary criticism theory and natural language processing technology. The results of sentiment analysis based on the Attention-BERT model show that the model performs well in the task of sentiment classification of residential space texts, with an accuracy of 83.53% and an F1 value of 80.05%, which is 10.49 percentage points higher than the 73.04% of the traditional LSTM model, verifying the synergistic advantages of the attention mechanism and the bidirectional coding structure.

The quantitative analysis at the lexical level shows that the residential spatial text has significant semantic density and diversity: the class-token-form-token ratio (TTR=12.01%) of the observation library is significantly higher than that of the reference library (7.27%), and the proportion of real words reaches 67.30% (compared with 52.95% in the reference library). The high frequency use of nouns (29.15% vs. 24.06%), verbs (20.91% vs. 18.06%), and adjectives (9.50% vs. 6.01%) directly reinforces the figurative nature of spatial depiction in relation to the dynamics of character behavior. The significant decrease in the percentage of dummy words (32.70% vs. 47.05%) further emphasizes the reliance of literary symbols on real word combinations.

Syntactic analysis shows that the distributional features of the quaternary word class clusters reveal the deeper association of space with power and identity: the NSPC class clusters (oriental noun dominant, frequency 98 vs. 25, LLR = 36.12) and VSPC class clusters (dynamic verb-led, frequency 115 vs. 48, LLR = 24.56) map social structures through physical boundaries and behavioral interactions; the negative structures (PRQV class clusters, frequency 37 vs. 6, LLR=40.18) and questioning tense (QSPQ class clusters, frequency 76 vs. 15, LLR=42.89), on

the other hand, reinforce the sense of repression and symbolic reflection, respectively, in the service of colonial modernity critique.

Fund Project

This paper is the final research outcome of the project of Social Science Development & Research Project in Hebei Province, titled “A Study on Translation Strategies for International Dissemination of Hebei’s Local Culture from the Perspective of Eco-Translatology” (Project Number: 202402134).

References

[1] Kwon, H. A., & Kim, S. (2019). Characteristics of residential space in response to changed lifestyles: Focusing on the characteristics of residents and the relationship between individual and family. sustainability, 11(7), 2006.
[2] Ultav, Z. T., Çağlar, T. N., & Drinkwater, S. B. D. (2016). Architectural literary analysis: Reading “The death of the Street” through Ballard’s literature and Trancik’s “Lost Space”. METU Journal of the Faculty of Architecture, 32(2).
[3] Caballero, R. (2014). Language, space and body: Sensing and construing built space through metaphor. Space, place and the discursive construction of identity, 107-134.
[4] Terzoglou, N. I. (2018). Architecture as meaningful language: Space, place and narrativity. Linguistics and literature studies, 6(3), 120-132.
[5] Yang, F., Xu, L., & Wang, J. (2025). Spatial Morphology of Urban Residential Space: A Complex Network Analysis Integrating Social and Physical Space. Sustainability, 17(5), 2327.
[6] Askarizad, R. (2018). Influence of socio-cultural factors on the formation of architectural spaces (Case study: Historical residential houses in Iran). Creative City Design, 1(1), 29-36.
[7] Sioli, A., & Kelsch, K. (2022). On literature and architecture: Imaginative representations of space. In Space and Language in Architectural Education (pp. 60-78). Routledge.
[8] Gazzaz, R., & Basnawi, A. (2024). Metaphor and architectural space in Emily Dickinson’s “I Dwell in Possibility”. The Explicator, 82(4), 193-198.
[9] Al-Azraki, A., & Al-Shamma, J. (2025). The Iraqi Home/Land under Siege: House as Metaphor in Abdul Razaq Al-Rubai's A Strange Bird on Our Roof. In Arabs, Politics, and Performance (pp. 31-45). Routledge.
[10] Akyıldız, E. C. (2021). An Architectural Reading of Franz Kafka s The Castle. In Architecture in Fictional Literature: Essays on Selected Works (pp. 1-17). Bentham Science Publishers.
[11] Hudson, B. J. (2021). Arnold Bennett, geography and architecture: A literary synthesis. Geoforum, 119, 94-101.
[12] Topolovská, T. (2022). Reading buildings: The textual turn of architecture as a parallel to the spatial turn in literary studies. Ars Aeterna, 14(1), 58-70.