

Publication Date: 20 August 2024

Archs Sci. (2024) Volume 74, Issue S2 Pages 50-59, Paper ID 2024s27.
<https://doi.org/10.62227/as/74s27>

An Empirical Study on the Effective Discourse Function of Teachers

Qiuyuan Tang^{1,*}

¹School of Foreign Languages, Inner Mongolia Minzu University, Tongliao, Inner Mongolia, 028000, China.

Corresponding authors: (e-mail: tangqiuyuan2022@163.com).

Abstract Based on the different categories of teachers' discourse in college English intensive reading classes, the characteristics of teachers' discourse in college English intensive reading classes can be categorized into two parts: formal characteristics and functional characteristics. Next, the Dirichlet distribution is applied to define and encode the formal definition of the task on the teacher discourse of college English intensive reading classes. Subsequently, it was further preprocessed and metrics and parameters were set to complete the construction of LDA topic segmentation model for effective discourse functions of teachers in college English intensive reading classes, and then the constructed model was used to conduct empirical research on teachers' discourse functions. The results show that the evolution of teachers' discourse data volume in college English intensive reading classes is divided into two life cycles: the first life cycle experienced the process of initial (56 in the first period)-explosion (669 in the 3rd period)-decline (49 in the 6th period), and the second life cycle took period 7 as the initial phase (77), which changed to an explosive phase (218) in period 9, and finally began to decline in period 10 (129), ending the second thematic discussion with initial - climax - decline. This study provides a better understanding of the quality and effectiveness of teacher discourse, improves teacher-student classroom relationships, improves teaching methods, and realizes an efficient English classroom.

Index Terms dirichlet distribution, LDA, metrics, teacher discourse functions

I. Introduction

Teachers' discourse plays a crucial role in organizing classroom teaching and facilitating students' language acquisition. As the intensive reading class for **non-English** majors constitutes an important foundational course in language teaching, the significance of teachers' discourse in this context cannot be overlooked. Teachers' discourse refers to all linguistic activities conducted by teachers in classroom teaching to impart knowledge and organize learning, permeating all aspects of classroom instruction, including explanation, questioning, feedback, error correction, and so on [1-3]. Research on teachers' classroom discourse has consistently been a key topic in linguistics, education, and psychology, with its findings not only confirming the importance of teachers' discourse but also providing effective theoretical guidance for teachers on how to employ discourse strategies to enhance the quality of classroom discourse. Different teacher discourse can evoke varying levels of student engagement and cognitive behavior in the classroom, ultimately leading to different learning outcomes [4-6]. The intensive reading class for **non-English** majors is a vital foundational course in language teaching, aiming at building a solid language foundation for students while developing their ability to communicate in English. Teachers' discourse in the intensive reading class

serves as an important source of language input for students and plays a modeling role for the target language [7-9]. Teaching contents can only achieve the desired teaching effect through the well-organized and effective delivery of teacher discourse. Given the significant role of teacher discourse in teaching non-English majors, and in order to enhance teaching efficiency, the author conducts research on teacher discourse in the intensive reading classroom through action research, with the aim of improving the validity of teacher discourse in this context, thereby promoting English language teaching and laying a solid foundation for students to develop their comprehensive language proficiency [10].

Literature [11] employs classroom discourse analysis to capture the characteristics of effective teacher frames in an intensive reading classroom based on CLI frames, aiming to inform the development of CLIL pedagogy in English as a foreign language education in China and other parts of the world. Literature [12] compares the discourse behaviors of two Chinese English teachers and conducts a quantitative analysis to reveal their different discourse characteristics. The highly rated teacher spoke most of the time, presented attractive information, and created a friendly atmosphere while maintaining a balance between text learning, language learning, and life experience sharing. Literature [13] con-

ducts a status quo analysis, finds and designs appropriate reading materials, creates and implements reading activities, and collects feedback. Learners' responses suggest that introducing specific materials may increase learner engagement and improve their comprehension. Specific, student-friendly reading materials foster positive attitudes and enhance high levels of comprehension. Literature [14] investigates teachers' understanding and use of digital discourse to support language arts instruction. Similar discrepancies were found between teachers' understanding of digital discourse and the prevalence of digital discourse practices in the classroom. There was a tension between teachers' motivation to use digital discourse and the complexity of students' implementation of it. Literature [15] analyzes the case of an ELL from a focal Asian background who encountered challenges when attempting to answer the teacher's questions and engage in classroom academic discourse about the earth sciences. The analysis showed that this ELL was not aware of the teacher's expectations regarding the intertextual connections and academic language needed to succeed in science tasks. Literature [16] discusses the role teacher feedback plays in extending conversations to encourage children to use language in meaningful ways. Research findings on teacher feedback are reviewed, and evidence-based, practical feedback suggestions are provided that are designed to support teachers and children in engaging in dialogue. Asking open-ended questions as part of a meaningful conversation extends the conversation and supports children's language development.

This paper focuses on the problems related to teachers' discourse in college English intensive reading classrooms and explores how to enhance the effectiveness of such discourse. With the support of big data technology, the research designs an investigation into the effective discourse functions of teachers. Based on distinct discourse categories and topics, the study initially categorizes the characteristics of teachers' discourse in college English intensive reading classrooms into two dimensions: formal characteristics and functional characteristics. It then examines the effectiveness of teachers' discourse in terms of the number of discourse instances, topic categories, questioning techniques, and feedback methods. The extracted discourse data were subsequently defined and coded using a combination of the Delicacy distribution and Gibbs sampling. Following this, further preprocessing operations, such as text segmentation and de-emphasis, were performed. Subsequently, metrics and parameters for discourse analysis were determined to complete the design of the LDA-based topic segmentation model. Ultimately, the constructed model is employed to empirically analyze the effective discourse functions of teachers in college English intensive reading classrooms.

II. A Study of Teachers' Effective Discourse Functions Supported by Big Data Technology

A. University english teachers' discourse and its characteristics

1) University english teachers' discourse

Teachers' discourse refers to the discourse employed by teachers who impart knowledge and organize teaching activities within the classroom setting. Teachers' discourse in the English classroom exhibits a dual characteristic: firstly, it serves as the source of the target language for students' second language acquisition; and secondly, it functions as a teaching tool for teachers to organize, manage, and guide the second language classroom. The significance of teacher discourse is manifested not only in classroom organization but also in the process of students' language acquisition. It is evident in classroom organization because teacher discourse directly influences the success or failure of the teacher's implementation of the teaching plan. Furthermore, it is manifested in the process of language acquisition because teacher discourse may constitute the most crucial linguistic input available to the learners.

2) Characteristics of university english teachers' discourse

We typically examine teacher discourse in the second language classroom from two perspectives: formal features and functional features. Formal features pertain to the linguistic characteristics of the teacher's discourse in the second language classroom, functioning as a 'simplified code' that offers comprehensible linguistic input to second language learners. Functional features, on the other hand, relate to the linguistic characteristics of teacher discourse in the second language classroom, serving as a pedagogical tool for organizing, managing, and guiding second language classroom activities.

- 1) Formal Characteristics of Teacher Discourse: Firstly, during language acquisition, the target language must constitute comprehensible input. The target language should be at the level of stage $i+1$, meaning it should be slightly more advanced than the learners' current language ability. In the English listening and speaking classroom, the author observed that all three teachers, consciously or unconsciously, made certain adjustments to the language forms in accordance with the comprehensible input theory. For instance, in terms of phonetics, they pronounced words clearly, spoke at a slow pace, and frequently paused. In terms of vocabulary, they endeavored to use basic words and minimized the use of slang or idioms. The syntactic structure was simple, primarily consisting of imperative and elliptical sentences. In terms of discourse organization, the overall structure was emphasized, and coherence was maintained as much as possible.
- 2) Functional characteristics of teachers' discourse: Two-way communication is more conducive to language learning than one-way communication. In two-way communication, when one party cannot understand the other, both parties can negotiate meaning by verifying understanding, confirming comprehension, and clarifying requests. This process triggers more repetitions and paraphrases of linguistic information, which can

increase comprehensible input, enhance information redundancy, and accelerate the second language acquisition process for learners. Therefore, teacher discourse in the classroom should promote communication and interaction between teachers and students. In this paper, the author examines the interactive function of teacher discourse from four aspects: the number of teacher utterances, the topic categories, the questioning methods, and the feedback methods.

Therefore, in the process of analyzing teacher discourse, the author examines the limitations of the language itself and seeks to understand the purpose of teachers' words in the classroom, with the aim of enhancing the practical value of teachers' language in the English classroom and ultimately enabling their words to have a greater impact in the classroom.

B. Theme segmentation model

1) Formalized definition of tasks

Before introducing our topic segmentation and classification model, we give formal definitions and symbolic notation for text topic segmentation and classification. Let D be a document with n sentences, $D = (s_1, s_2, \dots, s_n)$ and given K predefined topics whose textual representation is $U = (u_1, u_2, \dots, u_K)$, the task of text topic segmentation and classification is to divide D into semantically and thematically coherent segments (f_1, f_2, \dots, f_m) where $f_i = (a_i, b_i)$ contains the first position $a_i \in \{1, \dots, n\}$ of the i th segment and the topic label $b_i \in \{1, \dots, K\}$.

2) Delicacy distribution

Typically, the Dirichlet distribution is defined as:

$$Dir(\vec{p}|\vec{\alpha}) = \frac{\Gamma(\sum_{k=1}^K \alpha_k)}{\prod_{k=1}^K \Gamma(\alpha_k)} \prod_{k=1}^K p_k^{\alpha_k - 1}, \quad (1)$$

where $\vec{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_k)$ denotes the parameters of the Delikeray distribution and K denotes the dimension. $\Gamma(\cdot)$ is the Gamma function, which satisfies $\Gamma(n) = (n-1)!$. p_i satisfies $\sum_{k=1}^K p_i = 1$. $Dir(\vec{p}|\vec{\alpha})$ is generally denoted as $Dir(\vec{\alpha})$. When $\alpha_1 = \alpha_2 = \dots = \alpha_k$, it is denoted as $Dir(\alpha)$, which is called the symmetric Delikeray distribution. In the representation of the subject model, it is generally assumed that the prior distribution θ obeys the symmetric Dirichlet distribution $Dir(\alpha)$, denoted $\theta \sim Dir(\alpha)$. The conjugate distribution of a multinomial distribution is called the Dirichlet distribution, and when the likelihood function is a multinomial distribution and the prior distribution is the Dirichlet distribution, the posterior distribution is still the Dirichlet distribution. The Delikeray Process Mixed Modeling Assumption (DPMM) assumes that a mixed model produces a set of data and that the model contains several mixed components, an assumption that is essentially consistent with the idea of thematic modeling.

3) Gibbs sampling for LDA

The structure of the LDA topic model is shown in Figure 1, as can be seen from the document generation process of

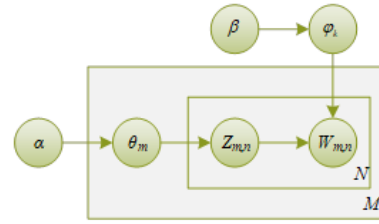


Figure 1: LDA theme model structure

the LDA model, the core problem of the topic model LDA lies in solving the values of the document-topic distribution $\vec{\theta}_1$ and the topic-word distribution $\vec{\phi}_k$. There are usually two typical algorithms for solving the parameter parsing, a deterministic approximation method, which first assumes a priori distribution, and then infers the priori distribution through the posterior distribution to derive the parameter estimates. The other is the Gibbs sampling method, which first constructs the Markov Chain and then derives the parameters of the posterior distribution with the extracted sample data. Since Gibbs sampling is to estimate the true probability distribution after iterative sampling, which is fast and convenient, this paper adopts Gibbs sampling algorithm to estimate the model parameters.

The Gibbs sampling method is a special case of the (MH) algorithm in which multiple joint probability distributions of random variables can be simulated and sampled when $alpha = 1$ and the distribution is multinomial. The idea of Monte Carlo Markov theory (MCMC) is that random data are generated by random sampling or simulation experiments and the posterior distributions are progressively approximated to numerical features such as expectation, integral, etc. in the probability model. The Metropolis-Hastings algorithm is one of the Markov theories. The main characteristic of a Markov chain is that the state machine is related only to the current state and not to the previous state. The chain reaches a smooth state after a sufficiently long running time. Gibbs sampling estimates the true probability distribution by means of iteration, which serves to eliminate the effects of parameter values when they are at initialization.

Gibbs sampling distribution is denoted as $p(\vec{v}|\vec{w})$, subject \vec{v} is the hidden variable, \vec{w} denotes the word variable, and the conditional distribution is $p(v_j = k|\vec{v}_{-j}, \vec{w})$. When the observation is $w_j = t$, the probability distribution is $p(v_j = k|\vec{v}_{-j}, \vec{w}) \propto p(v_j = k, w_j = t|\vec{v}_{-j}, \vec{w}_{-j})$ according to the Bayesian posterior formula, and in the LDA there is, again, a $\alpha \rightarrow \theta_i \rightarrow v_j, \beta \rightarrow \phi_k \rightarrow w_j$. Therefore, the posterior probability distributions of both $\vec{\theta}_1$ and $\vec{\phi}_k$ obey the Dirichlet distribution, that is:

$$p(\vec{\theta}_i|\vec{v}_{-j}, \vec{w}_{-j}) = Dir(\vec{n}_{i,-j} + \vec{\alpha}). \quad (2)$$

$$p(\vec{\phi}_k|\vec{v}_{-j}, \vec{w}_{-j}) = Dir(\vec{n}_{i,-j} + \vec{\beta}). \quad (3)$$

According to the Bayesian distribution there:

$$\begin{aligned} p(v_j = k | \vec{v}_{-j}, \vec{w}) &\propto p(v_j = k, w_j = t | \vec{v}_{-j}, \vec{w}_{-j}) \\ &= \int p(v_j = k, w_j = t, \vec{\theta}_i, \vec{\phi}_k | \vec{v}_{-j}, \vec{w}_{-j}) d_{\vec{\theta}_i} d_{\vec{\phi}_k} \\ &= E(\theta_{i,k}) \bullet E(\phi_{k,i}) \\ &= \hat{\theta}_{i,k} \bullet \hat{\phi}_{k,i}. \end{aligned} \quad (4)$$

From the parameter estimation of the Dirichlet distribution, the value of the document-topic distribution $\vec{\theta}_1$ is obtained:

$$\hat{\theta}_i = \frac{n_{i,-j}^k + \alpha_k}{\sum_k n_{i,-j}^k + \alpha_k}. \quad (5)$$

The value of subject-word distribution $\vec{\phi}_k$ is:

$$\hat{\phi}_{k,t} = \frac{n_{k,-j}^t + \beta_t}{\sum_t n_{k,-j}^t + \beta_t}. \quad (6)$$

The steps of Gibbs sampling algorithm are as follows:

Input: the corpus after word-splitting (in this paper, one document is one line of data), the number of topics K , and hyperparameters α and β . output: topic distribution of each document θ . word distribution of each topic ϕ . the topN topic words with the highest probability for each topic and the topic to which each word is assigned.

- Step 1: Randomly initialize all the words in the corpus.
- Step 2: Calculate the frequency size of each topic under different texts, the frequency of each word under each topic, and the topic distribution of other words except the current word $p(v_j | v_{-j}, w)$. Update the position $t = 0, 1, \dots, n$ and the topic of each word, estimate the probability that the current word belongs to each topic under the condition that the topics of the other words are known, and accordingly sample a new topic for the current word.
- Step 3: Iterate until the topic distribution $\theta_i = p(v_k | d_i)$ and word distribution $\phi_{i,j} = p(w_j | v_k)$ converge, ending the sampling process.

Gibbs sampling algorithm adopts stochastic gradient descent algorithm, which has a large error and needs to set the number of topics artificially in the LDA iteration process, which has a high time complexity. In addition, if the number of topics increases during the sampling process, its time complexity will also increase. Therefore, when the number of corpus set is large, only using LDA modeling algorithm for topic representation is still insufficient.

C. Application of theme segmentation modeling in discourse functional analysis

1) Experimental data collection

The experimental data collection is presented in Fig. 2. Due to the absence of a public test dataset for Chinese text segmentation, and considering that the research application focus of this paper is on classroom teaching texts, it was necessary to construct a high-quality text segmentation corpus in the teaching domain. The objective of segmenting teacher's lecture discourse is to divide the long lecture text into sub-theme

modules, where the semantic representation within each module is relatively complete, the theme is single, and the main idea and semantics of each sub-theme align with the teaching analysis sub-theme. After reviewing numerous texts online, we discovered that Baidu Encyclopedia's interpretations of college English texts possess the textual characteristics we require. Typically, an interpretation of a text is divided into content, background information, and full text appreciation, with the latter being described from various angles such as article structure, thoughts and feelings, writing techniques, etc., which align with the directions teachers take when explaining texts. In terms of text content, Baidu Encyclopedia meets our demands very well, and the encyclopedia's clearly labeled sub-topics can alleviate the burden of manual corpus segmentation. Another suitable corpus comes from college English text materials such as lesson plans and lecture notes. These articles also contain multi-angle analyses of college English texts, which are believed to be beneficial for theme extraction. The corpus involved in the experiment of this paper is primarily divided into four parts:

- 1) University English Texts Lesson Plans and Speaking Notes Texts: This part of the content was obtained by crawling text from lesson plans and lecture notes from high-quality teaching resource websites, and ultimately, after manual screening and filtering, we obtained 11,738 texts.
- 2) Baidu Encyclopedia College English Text Interpretation Texts: This part of the content was obtained by crawling texts from Baidu Encyclopedia based on the title directory of college English textbooks, subsequently labeling the data according to the subheadings of the texts, and ultimately acquiring 936 encyclopedic texts after manual screening and filtering.
- 3) Video Transcription of University English Courses: This part of the content comprises live video lessons that have been manually transcribed into text. Initially, research data was collected through offline classroom video recordings and crawling of online quality courses. Classroom observers were assigned to each experimental class to record teaching activities on video, while also compiling basic classroom information and lecture content, including class name, teacher's name, teacher's title, years of teaching experience, course category, time of collection, course name, and course format, etc. The data were gathered using a crawler program specifically designed to crawl the quality videos of "One Teacher One Excellent Lesson", along with relevant information such as the number of teacher's words, subject categories, questioning methods, and feedback methods. The collected classroom videos were then transcribed into dialog text format. This process began with using a file formatting tool to convert the video files into MP3 or WAV format. Subsequently, we utilized the speech transcription service provided by the KDDI platform and invoked its JAVA development interface to transcribe the

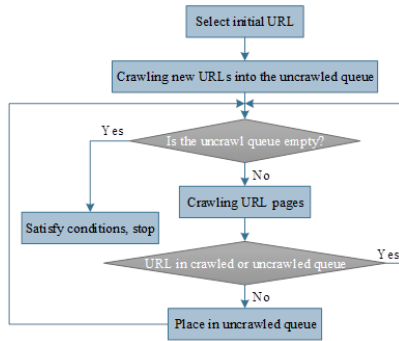


Figure 2: Experimental data acquisition

audio files into text files. Due to the potential errors in the platform's transcription results, manual corrections and annotations were made to part of the text, ultimately yielding 130 text files.

- 4) University English Wikipedia Dataset: Wikipedia's university English corpus is of high quality, extensive, and open. All entries are packaged monthly for users to download and use. Currently, Chinese Wikipedia has reached 860,000 entries.

2) Text pre-processing

- 1) Text tokenization: Words are the smallest independent units of meaning in a language system. Each word in English text is independent of each other, making word segmentation relatively simple compared with Chinese. In Chinese word segmentation, it is necessary to divide sentences without explicit separators into a series of independent words. There are basically three types of word segmentation methods commonly used in text mining. The first type is string-based matching, which involves comparing and matching a defined dictionary with the words in the text. If a match is successful, these words are added based on criteria such as "forward maximum match", "reverse maximum match", or "long word first". The second type is rule-based matching according to statistical methods. This first requires manually labeling the text in the corpus, with labeling based on statistical features or parts of speech. Then, the labeled text data is modeled, the parameters in the model are trained, the probability of the occurrence of each word in the corpus text is calculated, and the word segmentation result with the highest probability is taken as the final result. The advantage of this algorithm is that it solves the problem of word ambiguity, but the disadvantage is that it takes a lot of time in the manual annotation process and the word segmentation speed is too slow. The third type of word segmentation method is a recognition method based on knowledge understanding, and the core idea of the algorithm is to identify various linguistic information by simulating human understanding of sentences through machines.

- 2) Text de-duplication: The research object of text de-duplication is duplicate or near-duplicate text. For near-duplicate text, similarity calculation is commonly used to address such problems. The model generates vectors by counting the frequency of word occurrences in the text. In text de-duplication, multiple iterations are required to complete the comparison of texts with each other. Assuming there are Document 1 and Document 2, and a dictionary is constructed based on these two documents, the vectors generated for Document 1 and Document 2 are composed of the number of occurrences of words in the dictionary constructed from them.

Although VSM can count the frequency of occurrence of all words in the text, it also amplifies the characterization ability of some irrelevant words such as "is" and "the". Therefore, in the process of de-emphasis, the paper chooses another algorithm: TF-IDF. The basic idea of the TF-IDF model is that if a word appears more frequently in a document and less frequently in other documents, then the word has a higher level of importance to the document, and the topic of the document can be determined by this word. The advantage of this representation is that it removes a large number of words in the document that have nothing to do with its content, but at the same time retains the high-frequency phrases that can represent the main content of the document. In TF-IDF, TF is the frequency of occurrence of a word in the text, while IDF represents the sparseness of a word in all the corpus. The calculation formula is as follows:

$$TF = \frac{p_{ij}}{|w_j|}, \quad (7)$$

$$IDF = \log \frac{N}{N_i + 1}, \quad (8)$$

where, p_{ij} is the frequency size of word i in document j , w_j represents the total number of words in text j . N represents the number of all texts in the corpus, and N_i represents the number of texts that include the word i .

3) Metrics

In order to facilitate the comparison with similar algorithms, this paper adopts two metrics commonly used in text segmentation tasks P_k and WD. the proposed evaluation metric P_k is used to portray the probability of error that a certain two sentences do not fall in the same semantic paragraph or different semantic paragraphs in the correct segmentation result and the predicted segmentation result. P_k takes a value ranging from 0 to 1, with a lower score indicating a better segmentation result. P_k can be defined as:

$$P_k = P_{seg}P_{miss} + (1 - P_{seg})P_{false\ alarm}, \quad (9)$$

where P_{seg} is the probability that two sentences with a distance of k belong to different semantic passages. P_{miss} is the probability that a paragraph is missing from the segmentation result predicted by the algorithm. $P_{false\ miss}$ is the probability that a paragraph is added to the segmentation result predicted by the algorithm.

Although P_k can portray the error rate of text segmentation to a certain extent, the metric penalizes "negative errors" more than "positive errors", and is sensitive to changes in the size of segments, etc. To address these shortcomings, the method compares the difference in the number of correct and predicted segments within the window. To address these shortcomings, the method penalizes the difference in the number of segments by comparing the number of correct segments and predicted segments within the window, and penalizes the difference in the number of segments. the value of WD ranges from 0 to 1, and the lower the value is, the better the segmentation effect. It can be defined as:

$$\begin{aligned} & WindowDiff(ref, hyp) \\ &= \frac{1}{N-k} \sum_{i=1}^{N-k} (|b(ref_i, ref_{i+k}) - b(hyp_i, hyp_{i+k})| > 0), \end{aligned} \quad (10)$$

where k takes half of the average length of the sub-topic paragraph in the real segmentation result, $b(i, j)$ represents the number of segmentation boundaries between text units i and j , ref_i represents the position of the real segmentation boundary, hyp_i represents the position of the segmentation boundary predicted by the algorithm, and N represents the total number of text units in the document.

4) Parameterization

Gibbs sampling was used for the LDA model in the experiments, and in the process of topic modeling on the dataset, the values of hyperparameters α and β were first determined, and then the most appropriate values for the number of topics, T , were selected through multiple sets of experiments. The text takes empirical values for the two hyperparameters such that $\alpha = 50/T$, $\beta = 0.01$, T in 3, 5, 10, 15, 20, and 100 to run the algorithm and detect the changes in each evaluation metric.

According to the requirements of this experiment, the parameter settings for training the dataset with word2vec are shown in Table 1. In the process of model training using word2vec, we spliced the Chinese Wikipedia, all lesson plan data, Baidu encyclopedia data, and 50 transcribed texts to form a mixed dataset, which consisted of a total of 252,419 texts.

III. Empirical Analysis of English Teachers' Discourse under Thematic Modeling Theory

A. Analysis of the quantity and thematic categories of university english teachers' discourse

1) Quantitative analysis of university english teachers' discourse

The results of the analysis of the quantity of university English teachers' discourse are presented in Figure 3. A total of 1,480 pieces of eligible course discussion data were cleaned and obtained after the preliminary stages of text preprocessing and extraction of topic information. In this study, to determine the changes in the number of discussions in the course, a bar chart

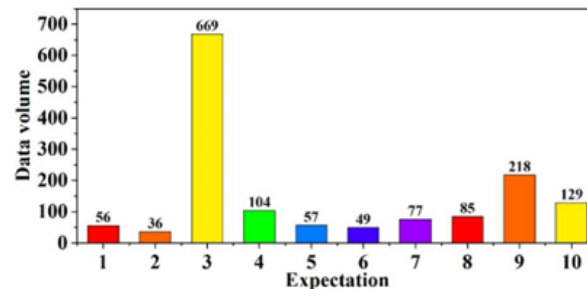


Figure 3: The results of the analysis of the number of English teachers' words in college

was drawn according to the trend of the data volume. It can be observed that the trend of the amount of teacher discourse data in the university English accuracy classroom exhibits a "bimodal" shape. This study divides the data evolution into two life cycles. The first life cycle starts in period 1 (0.56) and peaks in period 3, when the amount of data reaches 669 items. It then begins to decline in period 4 (0.104), and the decline period persists until period 6 (0.49). The first life cycle undergoes an initial-explosion-decline process. The second life cycle commences with an initial phase in period 7 (0.77), transitions to an outburst phase in period 9 (0.218), and ultimately begins to decline in period 10 (0.129), concluding the initial-explosion-decline of the second theme discussion. A comparison of the two life cycles with changing trends reveals that the "slope" of the life cycle of the second evolution has declined considerably compared to that of the first evolution.

2) Analysis of thematic categories of university English teachers' discourse

The topic probability distribution of university English teachers' discourse is presented in Table 2. In this paper, the 10 sets of topic results generated by the LDA topic model are sorted in ascending order according to the distribution probability of the topics. Additionally, the vocabulary within each topic is also sorted in ascending order based on the distribution probability of the vocabulary. The values are expressed as $0.276 > 0.128 > 0.092 > 0.082 > 0.046 > 0.032 > 0.026 > 0.018 > 0.015 > 0.008$, and the words with the top 10 distribution probabilities are extracted to express the thematic meanings.

Based on the observable information from the original data and in conjunction with the table, it can be seen that by observing the semantic information in topics 1 and 3, the words "teaching", "course", "harvest", "great benefit", and "guiding significance" in the semantic information are largely related to the feelings of learners after learning how to conduct teaching research. Teachers believe that the course has great guiding significance for their own teaching after learning the course content. In this theme, teachers express their gains and insights on knowledge during and after learning. Therefore, topics 1 and 3 are categorized as "learning perceptions".

By observing the semantic information in Themes 2, 8, and 10, it can be found that the terms "speech speed", "fast speed",

Hyperparameters	Parameter description	Value
Size	Word Vector Dimension	300
Window	Window size, i.e., maximum distance from word vector contexts	5
Min_count	Minimum word frequency of a word vector	1
Sg	Training algorithm, 1 for skip-gram, 0 for CBOW model	0
Hs	1 for stratified softmax, 0 for negative sampling	0

Table 1: Parameter setting

"difficulty", "rural teachers", "all English", "difficult to understand", and "professional terms" in the semantic information are all intended to reflect the teachers' response to the impact on their learning due to the lecturer's fast speaking speed and the challenging nature of the after-class examination. Rural teachers find it difficult to understand the terminology in English. The teaching and research explained in the class are perceived as difficult to practice, and suggestions for improvement are put forward. Under this theme, teachers expressed their belief that the curriculum was unreasonable and suggested improvements. Therefore, Themes 2, 8, and 10 are categorized as "improvement suggestions".

Upon observing the semantic information in Theme 5, it is evident that the words "front-line", "teacher", "further", "continuous progress", "occupation", "teaching craftsman", and "qualified" in the semantic information all reveal that teachers in this theme are aware of the gap between "teaching craftsmen" and truly qualified teachers. Therefore, Theme Five is categorized as "self-summary".

By observing the semantic information in topics 6, 7, and 9, it can be discerned that the words "eye-opening", "colorful", "featured", "information technology", "means", and "fascinating" within the semantic information are intended to convey the learners' comprehensive evaluation of the course content after participating in the course learning. In this theme, learners express that the course has been eye-opening, and various information technology means have facilitated extensive teaching research. Furthermore, learners emphasize the evaluation of course effectiveness, teaching methodologies, and other course-related content. Consequently, topics 6, 7, and 9 are categorized as "course evaluation".

In summary, the themes were categorized based on the degree of separation between them, and the original 20 sets of data results were consolidated into five distinct theme categories.

Next, the evolution of discourse themes among college teachers during the course learning is explored. The collected research data are analyzed according to the course opening cycle, revealing the trend of theme changes among college English teachers. The temporal changes of college English discourse topics are presented in Figure 4. The data indicates that in the first and tenth courses, the trend of topics discussed by primary and secondary school teachers exhibits a roughly "bimodal" pattern. In the initial stage, the volume of data discussed by primary and secondary school teachers in the first curriculum cycle was minimal, with each topic being relatively evenly distributed. The content related to the topic of "classroom evaluation" was slightly prominent during this

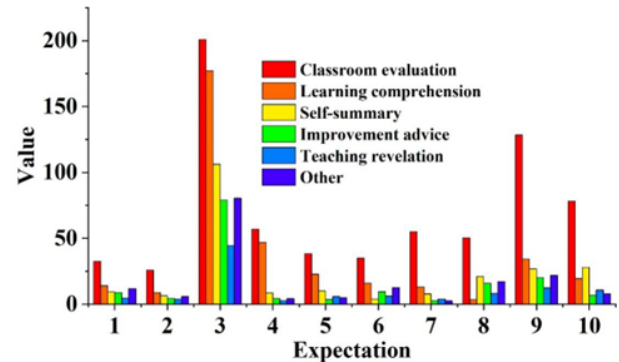


Figure 4: The change of the theme of college English discourse

period, which constituted the initial stage of discussion. In the subsequent explosive stage, the evolutionary line of the third course increased sharply, with the themes experiencing the largest increase being "classroom evaluation" (204) and "learning perception" (173). The theme of "self-summary" (112) also demonstrated relatively significant improvement. Primary and secondary school teachers actively participated in the discussion during the third period, but this high-frequency discussion was not sustained, resulting in the absence of a distinct "climax" stage in the first evolutionary life cycle. During the recession phase, the number of discussions on each topic in the 4th session plummeted and remained low until the 6th session, during which the discussions under each topic tended to be stable, with no significant change. Thus, the first topic discussion cycle of initial-outburst-recession concluded. By the 9th session, there was a minor peak surrounding the topic of "class discussions", which subsequently declined by the 10th session.

B. Empirical analysis of discourse questioning and feedback for college english teachers

1) Empirical analysis of university english teachers' discourse questioning style

Based on the thematic model, the coding of university English teachers' discourse questioning in the new knowledge teaching session of the intensive reading classroom was conducted. The results of the analysis, as presented in Table 3, reveal that the questioning in the new knowledge teaching session exhibits a generally similar situation to that of the introduction session. However, there are some notable differences. Firstly, there is an increase in closed questions (92.70%) and a

Theme	Theme vocabulary	The kind of species that belongs to
Theme 1(276)	Study, study, teaching, course, teacher, harvest, content, teacher, benefit greatly	Learning comprehension
Theme 2(0.128)	Speed, scholar, quick, single, exam, after-class, difficulty, feeling, vision, interest	Improvement advice
Theme 3(0.092)	Effective means, guidance, feeling, education, teaching research, teaching, guiding meaning, growth, large, research	Knowledge of the west
Theme 4(0.082)	Suitable, solution, teacher, foundation, learning, doubt, object, inspiration, development, reform	Teaching revelation
Theme 5(0.046)	Further, literacy, teachers, energy, guiding significance, career, teaching and writing, qualification, continuous progress, line	Self-summary
Theme 6(0.032)	Professional knowledge, open vision, promotion, development, rich and colorful, class, long-term, feature, academic research, open	Course evaluation
Theme 7(0.026)	Class, chapter, attention, whole, seriousness, goal, experiment, student, child, question	Course evaluation
Theme 8(0.018)	Experiments, control, and all aspects, and also, it is a pity, the countryside, the teacher, the meticulous, the whole English, the difficult to understand	Improvement advice
Theme 9(0.018)	Various kinds of information technology, means, use good, experience, just, appropriate, fascinating, multiple forms, solutions	Course evaluation
Theme 10(0.008)	Key, professional terms, absorption, progress, a little, practicality, difficulty, new, deep, detailed	Improvement advice

Table 2: The probability distribution of English teacher discourse

Question quality	Problem type	Closed	Proportion	Proportion	Open
Ask		541	39.43%	33	30.56%
Reduplication		114	8.31%	11	10.19%
Omission		291	21.21%	29	26.85%
Replenish		84	6.12%	8	7.41%
Hint		174	12.68%	14	12.96%
Evaluation		168	12.24%	13	12.04%
Total		1372	100.00%	108	100.00%
Proportion		92.70%		7.30%	

Table 3: An empirical analysis of the question of English teachers in college

corresponding decrease in open questions (7.30%). Secondly, within the open questions, there is an increase in elliptical questions (21.21%). Both of these trends can be attributed to the overall decrease in open-ended questions, as elliptical questioning essentially involves repetition of elicitation questions. Therefore, open-ended questions significantly decrease in the new knowledge teaching session. This may be because the new knowledge teaching session is primarily focused on explaining knowledge. The classroom introduction, on the other hand, may establish a scenario that aligns with students' life experiences and mathematical knowledge to stimulate their interest in learning new knowledge and awaken their awareness of the college English knowledge necessary for learning new content. In such an introduction, open-ended questions are feasible and highly suitable. However, for the abstract mathematical knowledge explained in the new knowledge teaching session, the increase in closed questions is inevitable, as this session is more geared towards students mastering the new knowledge within a specified timeframe. Compared to the uncertainty of open questions, closed questions are easier for teachers to use and grasp when explaining new knowledge.

2) Empirical analysis of university english teachers' discourse feedback styles

Firstly, the LDA topic model was employed to classify university English teacher feedback into two primary categories: one for the feedback provided by teachers when students answered their questions correctly, and the other for the feedback given

when students did not answer or answered incorrectly. These two types of feedback were coded as separate categories.

In one category of coding, feedback from college English teachers was classified into 11 subcategories, which were:

- 1) repetition,
- 2) board use,
- 3) weak verbal feedback,
- 4) strong verbal feedback,
- 5) follow-up questions,
- 6) teacher explanation,
- 7) repetition + board use,
- 8) repetition + weak verbal feedback,
- 9) repetition + strong verbal feedback,
- 10) letting the whole class evaluate, and
- 11) no feedback.

In the second type of coding, college English teacher feedback was classified into 10 categories, namely:

- 1) prompts,
- 2) repetitions,
- 3) corrections,
- 4) follow-up questions,
- 5) rhetorical questions,
- 6) repetition of the question,
- 7) reported errors,
- 8) teacher explanations,
- 9) teacher explanations combined with questions, and
- 10) and whole-class evaluations.

The results of the one-category coding analysis are presented in Table 4. The data reveal that, in comparison to the classroom introduction session, the teachers' one-category feedback in the new knowledge teaching session exhibits several differences. Firstly, eleven types of feedback modalities are represented, indicating that as the situation becomes more complex, the teacher's feedback is likely to encompass a greater variety of modalities. This is because the new knowledge teaching session involves significantly more time and communication with students, making it inherently more complicated than the classroom introduction. This complexity is reflected in the actual classroom teaching, where students'

Coding	Quantity	Proportion
1	343	23.18%
2	74	5.00%
3	184	12.43%
4	119	8.04%
5	261	17.64%
6	241	16.28%
7	49	3.31%
8	8	0.54%
9	29	1.96%
10	70	4.73%
11	102	6.89%

Table 4: A class of coding analysis results

responses become more unpredictable. As a result, teachers need to improvise when the process is derailed by unexpected situations. Therefore, the feedback employed by teachers in the new knowledge teaching session is more diversified. Secondly, feedback in categories 5 (261, 17.64%) and 6 (241, 16.28%) outnumbered feedback in categories 3 (184, 12.43%) and 4 (119, 8.04%), with categories 5 and 6 tending to be equal. Categories 5 and 6 consisted of follow-up questions and teacher explanations, respectively, while categories 3 and 4 consisted of weak verbal feedback and strong verbal feedback. The increase in follow-up questions and teacher explanations suggests that teachers are oriented towards ensuring students master new knowledge during the new knowledge teaching session. Follow-up questions prompt students to delve deeper into a certain knowledge point, enabling them to better grasp the new knowledge. Teacher explanations, on the other hand, allow students to re-examine what they have just learned, similarly facilitating better understanding of the new knowledge. Thirdly, there was a reduction in "no feedback", which refers to the teacher directly moving into the next section without responding to students' answers. The reduction of this type of feedback, along with the increase in categories 5 and 6, can be attributed to the teacher's focus on guiding students to master new knowledge and reinforcing learning through targeted feedback.

The results of the analysis of the two types of coding are shown in Table 5, which shows that the reduction of 3 (75, 5.07%) feedback methods, 3 types of feedback is the teacher directly modify the student's answer, if the feedback method of asking questions about others can still allow students to construct knowledge on their own, then the direct modification of the students deprived of the right to think, for the new knowledge teaching session is obviously unfavorable. And also for teachers, the use of revision appears mainly in simple questions, involving errors in content related to new knowledge, teachers rarely hang to not make indirect revision of the feedback method. 8 (50, 3.38%), 9 (45, 3.04%), 10 (42, 2.84%) types of feedback appear, that is, the teacher explains, the teacher explains and asks questions to his students as well as the whole class evaluates the student's answers. The type 1 prompts the most feedback, followed by the 2 class, which appears at 38.38% and 32.30%. Teacher explanations appeared more often because the answers of the students'

Coding	Quantity	Proportion
1	568	38.38%
2	478	32.30%
3	75	5.07%
4	61	4.12%
5	53	3.58%
6	60	4.05%
7	48	3.24%
8	50	3.38%
9	45	3.04%
10	42	2.84%

Table 5: Second class coding analysis results

responses were not within the scope of the discussion of the teaching at this stage, so the teacher tended to explain them, as in the example given during the coding. The way the whole class is evaluated is that the teacher wants to get a sense of how well the students have grasped the question; if the student answers wrongly and the whole class agrees with the student's answer, it means that the majority of the students do not yet have a good grasp of the knowledge involved in the question. Instead, only the student's question was asked, which allowed the teacher to make better choices about subsequent teaching behaviors.

IV. Conclusion

Teachers' discourse in college English intensive reading classrooms holds special significance and plays a dual role. This study begins with the function of teachers' discourse and adopts the LDA topic model to empirically analyze the effectiveness of discourse in college English intensive reading classrooms.

- Through the analysis of the number of discourses and subject categories of college English teachers, it is evident that the trend in the discourse data volume of college English precision classroom teachers exhibits a "bimodal shape". The first lifecycle begins in the first period (56), reaches a peak in the third period (669), and subsequently declines from the fourth period (104) onwards, persisting until the sixth period (49). The second lifecycle initiates in the seventh period (77), experiences an explosive stage in the ninth period (218), and ultimately enters a recession phase in the tenth period (129), concluding the second theme discussion with a beginning-climax-recession pattern. During the explosive stage, the evolutionary line of the third course increases sharply, with the themes of "classroom evaluation" (204) and "learning perception" (173) experiencing the largest growth among all theme categories, while the theme of "self-summary" (112) also exhibits notable improvement. Overall, the analysis of the number of college English teachers' discourses and topic categories reveals the significance of college English precision classroom teachers' discourse in the teaching process.
- Based on the coding of teacher discourse questioning and feedback in the intensive reading classroom of college English according to the thematic model, the question-

ing in the new knowledge delivery session exhibited a largely similar situation, with most of the questions asked by the teacher being closed questions (92.70%), which were relatively simple in comparison to the open questions (7.03%) that focused on textbook-based knowledge points. In addition, the feedback in category 5 (261, 17.64%) and category 6 (241, 16.28%) exceeded that of category 3 (184, 12.43%) and category 4 (119, 8.04%), and the feedback in categories 5 and 6 tended to be equal. This indicates that the teacher was oriented towards enabling students to master the new knowledge during the new knowledge teaching session, as the follow-up questions were continuations for the students, and this type of feedback was a means by which the teacher sought to facilitate a deeper understanding of a certain knowledge point, enabling students to better grasp the new knowledge.

References

- [1] Correa, D., & Flórez, I. (2022). Representations of foreign English teachers in Colombian newspapers: A critical discourse analysis. *Revista signos. Estudios de Lingüística*, 55(108), 6-36.
- [2] Banister, C. (2023). Decolonizing business English: exploring classroom ideologies. *ELT Journal*, 77(2), 166-175.
- [3] Ortega, Y. (2021). Transformative pedagogies for English teaching: Teachers and students building social justice together. *Applied Linguistics*, 42(6), 1144-1152.
- [4] Chang, S. H. H., & Huang, L. F. (2018). The Use of OK by Native and Non-Native Teachers in Bilingual Classrooms: Mandarin, English and Code-Switching. *Concentric: Studies in Linguistics*, 44(2), 111-135.
- [5] Lewis, M. A., & Zisselsberger, M. G. (2019). Scaffolding and inequitable participation in linguistically diverse book clubs. *Reading Research Quarterly*, 54(2), 167-186.
- [6] Sosiak, M. (2017). Professionally oriented discourse and status-role relations in the context of foreign language teacher training <http://ir.dspu.edu.ua/jspui/bitstream/123456789/1035/1/%D1%81%D0%BE%D1%81%D1%8F%D0%BA.pdf>.
- [7] Bannister-Tyrrell, M., & Clary, D. (2017). Metacognition in the English classroom: Reflections of middle years teachers navigating the Australian curriculum: English. *English in Australia*, 52(3), 63-72.
- [8] Flowerdew, L. (2019). Corpus Linguistics for English Teachers: New Tools, Online Resources, and Classroom Activities Eric Friginal. New York, NY: Routledge, 2018. Pp. xvi+ 349. *TESOL Quarterly*, 53(2), 594-596.
- [9] İnan Karagül, B. A. N. U., Aykut, C. A. N. S. U., & Ulu, B. (2020). Analysis of Articles and Theses on Discourse of English as a Foreign Language Classrooms Published Between the Years of 2010 and 2019 in Turkey. *Turkish Studies-educational Sciences*, 15(1), 1-13.
- [10] Douglas, K. (2018). Queer pedagogies and the university english classroom: Reflections on teaching EM Forster's the longest journey. *English in Australia*, 53(2), 78-82.
- [11] Li, D., & Zhang, L. (2022). Exploring teacher scaffolding in a CLIL-framed EFL intensive reading class: A classroom discourse analysis approach. *Language Teaching Research*, 26(3), 333-360.
- [12] John, D., & Devi, G. S. (2021). Designing STEM-Specific Student-Friendly Reading Content for the Engineering English Classroom. *IEEE Transactions on Professional Communication*, 64(4), 444-455.
- [13] Yang, X., & Tao, X. (2018). Comparing discourse behaviors of a high-rated and a low-rated Chinese EFL teacher: A systemic functional perspective. *The Modern Language Journal*, 102(3), 594-610.
- [14] Higgs, J. (2020). Digital discourse in classrooms: Language arts teachers' reported perceptions and implementation. *Research in the Teaching of English*, 55(1), 32-55.
- [15] Lan, S. W., & de Oliveira, L. C. (2019). English language learners' participation in the discourse of a multilingual science classroom. *International Journal of Science Education*, 41(9), 1246-1270.
- [16] Wasik, B. A., Farrow, J., & Hindman, A. H. (2022). More than "good job!": the critical role of teacher feedback in classroom discourse and language development. *The Reading Teacher*, 75(6), 733-738.