Journal of Combinatorial Mathematics
and Combinatorial Computing
www.combinatorialpress.com/jcmcc

Combinatorial Press

An Innovative Teaching Model of Japanese Language in Colleges and Universities Based on Natural Language Processing Technology

Lixia Cui[1,✉], Lixian Xu[2]

[1] *Xi'an Fanyi University, Xi'an, Shaanxi, 710105, China*

[2] *Sanechips Technology Co., Ltd., Shenzhen, Guangdong, 518055, China*

ABSTRACT

The traditional Japanese language teaching mode in colleges and universities has been unable to meet the requirements of Japanese language majors in various industries, and colleges and universities should use certain methods to carry out a reasonable reform of the teaching mode of Japanese language majors. Firstly, an error correction model based on UniLM model framework is proposed, using natural language processing technology to extract features, and fine-tuning training for the model after initialization. The model framework based on UniLM+CRF and the seq2seq model framework based on UniLM are built to realize the Japanese text grammar error annotation task and the Japanese text grammar error correction task respectively. Then a multi-task learning error correction method is proposed to integrate the grammar error labeling task and the grammar error correction task, so as to improve the accuracy of the error correction model. Finally, a specific Japanese grammar error correction system architecture is designed, a Japanese language knowledge base is established, and utterance synthesis rules are formulated to realize the innovative teaching of Japanese language in colleges and universities. The average grammatical error correction precision, recall, and F1 value of the model in this paper reached a good level in the students' Japanese composition correction. The error between the average score of teacher correction and the average score of model correction is only 0.19 points, and the related experiments show that the innovative teaching model studied in this paper can effectively improve students' mastery of Japanese syntactic ability. The above data illustrate that the Japanese error correction system based on UmiLM framework designed in this paper has certain application value and can realize the innovation of Japanese language teaching mode.

Keywords: Natural Language Processing, Japanese Error Correction Model, UniLM, seq2seq model, Innovative Japanese Language Teaching

✉ Corresponding author.
E-mail address: cuilixia1@163.com (L. Cui).

Received 15 January 2024; Revised 25 March 2024; Accepted 30 November 2024; Published Online 15 April 2025.
DOI: 10.61091/jcmcc127a-211

1. Introduction

Natural language processing technology, as a branch of artificial intelligence technology, focuses on solving a variety of real-world problems through algorithms and data, where algorithms are clear instructions that allow computers to solve problems. The wide application of natural language processing technology has become a key aspect of today's data processing field [1-3]. When dealing with text data in different languages, it has become particularly important to utilize natural language processing techniques to process them. Processing multilingual text data requires multiple steps such as language recognition, word segmentation, lexical annotation, named entity recognition, translation and alignment, sentiment analysis, and topic modeling [4-7]. Using these natural language processing techniques, we can better process and understand multilingual text data, which plays an important role in various fields, and Japanese language teaching in colleges and universities is no exception [8-9].

With the continuous development of China's higher education, foreign language teaching as a major component of the talent cultivation model in colleges and universities is also developing continuously. In recent years, while English-based foreign language education has made great progress in colleges and universities all over the country, Japanese language majors have risen to the forefront, and have become a popular major second only to English majors in terms of the size of the student body [10-13]. Although the classroom teaching of Japanese language majors has also made certain development, in many institutions of higher education, Japanese language majors start from zero. Although the education of Japanese language majors in colleges and universities has made great progress in terms of faculty allocation and teaching hardware facilities [14-16]. However, after all, due to the limitations of objective reasons, the classroom teaching mode of Japanese language majors, especially the lower grades of Japanese language majors, still remains in the relatively traditional teaching mode [17-18].

Literature [19] aims to investigate Japanese text sentiment classification by combining intelligent machine algorithms. The TF-IDF algorithm was combined with SVM to create a Japanese text sentiment classification model, and controlled experiments were designed to analyze the performance of the model. The comparative analysis shows that the research model has good comprehensive performance and can meet the needs of the sentiment classification system. Literature [20] outlines the development of a computer-assisted foreign language learning grammar knowledge base for learning Japanese sentence patterns using NLP and machine learning techniques. A combination of CRF machine learning algorithms and manual rules was also proposed. The experimental results point out that the program for designing and building a Japanese grammar knowledge base using NLP and machine learning techniques is effective. Literature [21] pointed out the problems in Japanese language teaching, stating that a single teaching method and boring teaching content cause students to lose interest in learning Japanese. It also introduced the application of seminar teaching method in Japanese language teaching, and the experimental results verified the effectiveness of the method and its help to improve students' learning and thinking ability. Literature [22] describes the error analysis system used to help the language learning process. It uses several natural language processing tools to evaluate the grammar of Japanese input and can analyze a wide range of grammatical error types. By collecting a certain number of sentences for systematic evaluation, it shows good accuracy and obtains a good evaluation of the user experience. Literature [23] emphasizes the application of AI natural language processing technology in Japanese language teaching, which is actually a language processing technology that combines computer science and AI. The application of AI natural language processing technology in Japanese language teaching is examined with a view to informing the research on educational technology for Japanese language teachers. Literature [24] built a Japanese translation teaching corpus based on a bilingual non-parallel data model, and based on the training in Moses proved that the model has good translation retrieval performance. And it also shows excellent performance and efficiency in assisting Japanese translation teaching.

This paper introduces the theoretical foundations of natural language processing and deep learning, and in this way constructs a Japanese grammar error correction model based on the UniLM framework. The vector embedding layer of the UniLM model consists of word vectors, position vectors, and distinguished sentence category vectors. The structure of the grammar error annotation model is UniLM+CRF, that is, a CRF layer is connected to UniLM, and the CRF layer decodes the predicted sequence of the model to find the optimal path as the final output sequence, while the grammar error correction task model adopts the end-to-end seq2seq architecture. Finally, these two single-tasks are trained jointly by multi-task learning, and the losses of the two-task model are superimposed for parameter optimization to improve the accuracy of the model. Methods such as comparative experiments were applied to validate the performance and effectiveness of the models, and intergroup and intragroup pre and post-test analyses were also applied to verify the positive effect of the innovative Japanese language teaching model designed in this paper on students' syntactic error rates.

2. Natural language processing techniques and deep learning

Natural Language Processing (NLP) is an important area of research that seeks to explore how computer technology can be utilized to learn, understand, and apply human discourse, as well as how to achieve effective communication between humans and computers, thereby increasing people's efficiency and productivity.

2.1. Segmentation techniques

Rule-based participles are centered on the idea of matching a given string of text with words in a dictionary in order to separate them.

In order to obtain a more accurate word separation, a combination of forward and reverse maximum matching methods together, also known as bidirectional matching method, can be adopted.

With the two-way maximum matching method, the words with the least word cuts can be compared in order to get the optimal segmentation result.

Statistical word splitting is usually accomplished in two steps:

(1) Establishing a statistical language model.

(2) By applying statistical learning algorithms such as Implicit Markov and Conditional Random Fields for sentence word segmentation, probability calculations are performed on the obtained segmentation results to arrive at the optimal segmentation method.

The language model is to compute the probability distribution $P(w_1, w_2, \cdots, w_m)$ of strings of length m and measure their probability values using the chaining method as shown in the following equation:

$$\begin{aligned} &P(w_1, w_2, \cdots, w_m) = P(w_1)P(w_2 \mid w_1)P(w_3 \mid w_1, w_2) \\ &\cdots P(w_i \mid w_1, w_2, \cdots, w_{i-1}) \cdots P(w_m \mid w_1, w_2, \cdots, w_{m-1}) \end{aligned} \tag{1}$$

The calculation of $P(w_i \mid w_1, w_2, \cdots, w_{i-1})$ can be simplified to:

$$P(w_i \mid w_1, w_2, \cdots, w_{i-1}) \approx P(w_i \mid w_{i-(n-1)}, \cdots, w_{i-1}) \tag{2}$$

When $n = 1$, we call this language model a unitary language model, and the probability of entering a complete sentence into the model is then $P(w_1, w_2, \cdots, w_m) = P(w_1)P(w_2) \cdots P(w_m)$.

When $n = 2$, this model is called a binary model and Eq. becomes $P(w_i \mid w_1, w_2, \cdots, w_{i-1}) = P(w_i \mid w_{i-1})$.

When $n = 3$, the ternary model, Eq. becomes $P(w_i \mid w_1, w_2, \cdots, w_{i-1}) = P(w_i \mid w_{i-2}, w_{i-1})$.

The probability of the n-tuple condition is estimated by scaling the frequency counts as shown in equation (3):

$$P(w_i \mid w_{i-(n-1)}, \cdots, w_{i-1}) = \frac{count(w_{i-(n-1)}, \cdots, w_{i-1}, w_i)}{count(w_{i-(n-1)}, \cdots, w_{i-1})} \tag{3}$$

where $count(w_{i-(n-1)}, \cdots, w_{i-1})$ represents the total number of occurrences of word $w_i - (n-1), \cdots, w_{i-1}$ in the corpus.

The main idea of the Hidden Markov Model (HMM) is that each character plays a certain "role" when it functions in a word, and is therefore labeled as an HMM after word separation.

Abstracting the central idea of the above HMM model, we get: the input data is denoted as $\lambda = \lambda_1 \lambda_2 \cdots \lambda_n$, n is the length of the sequence, λ_i is the character, and $o = o_1 o_2 \cdots o_n$ is the label of the output, then the final expected result should be:

$$\max = \max P(o_1 o_2 \cdots o_n \mid \lambda_1 \lambda_2 \cdots \lambda_n) \tag{4}$$

Given that the conditional probability $P(o \mid \lambda)$ of 2n variables is not stable, it seems impossible to make an accurate count of them. For this reason, we propose the observation independence hypothesis, which is calculated as follows:

$$P(o_1 o_2 \cdots o_n \mid \lambda_1 \lambda_2 \cdots \lambda_n) = P(o_1 \mid \lambda_1) P(o_2 \mid \lambda_2) \cdots P(o_n \mid \lambda_n) \tag{5}$$

HMM is an effective solution to solve Problem $P(o \mid \lambda)$ by using Bayesian formulation, which leads to more accurate results and hence better solutions.

$$P(o \mid \lambda) = \frac{P(o, \lambda)}{P(\lambda)} = \frac{P(\lambda \mid o) \mid P(o)}{P(\lambda)} \tag{6}$$

λ refers to the input data into the algorithm, and $P(\lambda)$ can be ignored as a result of the computation, so maximizing $P(o \mid \lambda)$ is equivalent to maximizing $P(\lambda \mid o) P(o)$.

Making Markov assumptions for $P(\lambda \mid o) P(o)$ yields:

$$P(\lambda \mid o) = P(\lambda_1 \mid o_1) P(\lambda_2 \mid o_2) \cdots P(\lambda_n \mid o_n) \tag{7}$$

Meanwhile, for $P(o)$ there:

$$P(o) = P(o_1) P(o_2 \mid o_1) P(o_3 \mid o_1, o_2) \cdots P(o_n \mid o_1, o_2, \cdots, o_{n-1}) \tag{8}$$

The HMM gives a new hypothesis, the chi-squared Markov hypothesis, which states that the content of the model's output at a given moment is closely related only to the content of the output at the previous moment, and thus, it can be derived:

$$P(o) = P(o_1) P(o_2 \mid o_1) P(o_3 \mid o_2) \cdots P(o_n \mid o_{n-1}) \tag{9}$$

And so:

$$P(\lambda \mid o) P(o) \sim P(\lambda_1 \mid o_1) P(o_2 \mid o_1) P(\lambda_2 \mid o_2) P(o_3 \mid o_2) \cdots P(o_n \mid o_{n-1}) P(\lambda_n \mid o_n) \tag{10}$$

By setting certain $P(o_k \mid o_{k-1}) = 0$, the appearance of unreasonable combinations such as BBB and EM can be effectively avoided.

In HMM, Veterbi's algorithm is an effective dynamic planning method, which can effectively plan out the optimal path from the initial node to point $o_i - 1$ and through o_i.

2.2. Word Embedding Techniques

Neural network techniques can be used to represent a variety of different expressions that are often referred to as word vectors, word embeddings, or distributed expressions.

2.2.1. Word Vector. In NLP tasks, the first step is usually to create a kind of word list library and number all the words in order. This approach is called One-hot representation. As the vocabulary increases, the dimensionality of the bag-of-words model will appear larger and larger and the matrix will appear very sparse. This increase in dimensionality leads to a significant consumption of computational resources. To better characterize the connection between the target word and the context, we must extract the relevant context data from the word vector.

2.2.2. Word2vec. Word2vec is an open source word vector modeling software [25] which uses Neural Network Language Model (NNLM) computation.

Word2vec neural network consists of input, projection, hidden and output layers. The main methods of computation are: first, reflecting each phrase onto a feature vector, second, using this feature vector to represent the probability assignment of the phrase, and finally, using the phrase data to reason about the parameters of the feature vector and the probability assignment.

Word2vec is able to represent words as real-valued vectors and contains both CBOW and Skip-Gram models.

2.2.3. ELMo. The ELMo model [26] is a pre-trained language model for generating word vectors, which is based on the input textual information (a sequence of words), and utilizes a bi-directional LSTM model to predict the next word in the forward direction and the next word in the reverse direction of the textual sequence, respectively, in order to train a dynamic word vector model.

2.3. Deep learning

Deep learning is a technique that combines linear and nonlinear components and is based on Artificial Neural Networks (ANN).

Sequence-to-sequence model is an end-to-end structured model that can handle time series related tasks by computationally realizing the data of one sequence from the data of another sequence. For example, given sequence data $x=x_1x_2\cdots x_s$, after the computation of seq2seq model, another sequence will be obtained $y=y_1y_2\cdots y_T$. Seq2seq model [27] has a strong capability of long range feature extraction compared to recurrent neural network model, and it is more capable in the dependency extraction of long range words.

The Seq2seq model consists of two units: an encoder and a decoder. The encoder is responsible for collecting and processing data $x=x_1x_2\cdots x_s$ from the outside world and transforming it into vectors C with specific meanings to better understand and represent the data. When time step t occurs, the encoder will calculate the hidden state h_i of the current time step based on the input x_i and the hidden state h_{i-1} of the previous time step, and the specific operation steps can be referred to Eq. (11):

$$h_t=f(h_{t-1},x_t) \tag{11}$$

The encoder computes the semantic vector C based on the hidden state of each time step after the computation is completed, and the computation process is shown in Equation (12), where q is a customized function:

$$C=q(h_1,h_2,\cdots,h_s) \tag{12}$$

Decoder is responsible for decoding, it takes the semantic vector C output from Encoder as input and the output y_{r-1} from the previous time step as input, and repeats this step until it encounters the termination flag, and finally realizes the decoding of the target sequence $y=y_1y_2\cdots y_T$. When time step t arrives, the Decoder will measure the hidden state of the time step h'_{t-1} based on the input y_{t-1},

the semantic vector C and the hidden state of the previous time step. Step's hidden state $h_t^{'}$, the specific calculation process can refer to equation (13):

$$h_t^{'} = f^{'}(h_{t-1}^{'}, y_{t-1}, C) \tag{13}$$

The specific calculation procedure for output y_t of time step t is shown in equation (14):

$$P(y_t \mid y_{t-1}, y_{t-2}, \cdots, y_1, C) = g(h_t^{'}, y_{t-1}, C) \tag{14}$$

The attention mechanism in neural network can effectively allocate computational resources to solve the information overload problem.

Firstly, the data in Source is converted into the form of key-value pairs, then the weight coefficients between Query and Key are calculated according to the correlation function and added to Value, and finally an attentional value score sequence of the word sequence can be obtained by weighted summation:

$$f(Q, K_i) = \begin{cases} Q^T K_i & dot \\ Q^T W_a K_i & general \\ W_q [Q : K_i] & concat \\ v_a^T \tanh(W_a Q + U_a K_i) & preceptron \\ \dfrac{Q^T K_i}{\sqrt{d}} & scale \end{cases} \tag{15}$$

$$a_i = soft\max(f(Q, K_i)) = \frac{\exp(f(Q, K_i))}{\sum_j \exp(f(Q, K_i))} \tag{16}$$

$$Attention(Q, K, V) = \sum_i a_i V_i \tag{17}$$

Attention values are calculated using a weighted summation approach, which is known as a soft attention mechanism.

3. Japanese grammar error correction model based on the UniLM framework

3.1. The UniLM model

Both single-task and multi-task models use the UniLM model as the Encoder side to extract features and then fine-tune them based on the pre-trained model parameters. In this paper, the Japanese pre-training model RoBERTa and the open source library-transformers provided by Hugging face are used.

The vector embedding layer is a high latitude vector to characterize the text sequences input to the model.The vector embedding layer of the UniLM model consists of three parts: word vectors, position vectors, and vectors distinguishing sentence categories.

Token embeddings represent the word sense features of the input sequence, if the input sequence consists of 2 or more sentences, segment embeddings are needed to indicate whether the current token belongs to the left sentence or the right sentence.Position embeddings represent the positional features of the token.

The input sequence is mapped by word dictionary to get one-hot encoding, and then mapped in the pre-trained word vector matrix to get the token embeddings of the input sequence.Since the input sequence of the UniLM model consists of two sentences, it is necessary to generate an encoding consisting of 0 and 1, with 0 indicating that the current token belongs to sentence 1 and 1 indicating that the current token belongs to sentence 2 , also mapped in the pre-trained sentence category vector matrix. The one-hot positional encoding of the input sequence is the positional number where each

token in the sequence is located, e.g., if the length of the sequence is 10, a one-hot encoding of [0, 1, ..., 9] is generated. Similarly the position vectors of the input sequence are obtained by mapping through the pre-trained position vector matrix. All three vectors are involved in fine-tuning and learning during model training.

The multi-head attention layer of the UniLM model is the same as that used in the general BERT model, which is self-attentive. The difference is that the attention mask matrix in UniLM not only masks the filled part, but also designs a special mask matrix for masking operation according to the model task and the type of input. During the training process, it is ensured that the attention of each token of the input sequence does not fuse the information of future tensors, and also ensures that the attention of each token pays full attention to the known information.

3.2. Japanese grammar error labeling task

The training of neural networks is highly computationally complex, and in order to reduce the time cost of training and speed up the convergence of the network, normalization is often performed, and the method used here is layer normalization.

Inside the attention module of the model, there is a residual connection behind the sub-layer self-attention layer and the fully connected feedback layer, which will sum up the inputs and outputs of the network layers, effectively solving the problems of gradient disappearance, gradient explosion and overfitting that arise with the deepening of the first depth network layers.

The syntax error labeling model is based on UniLM+CRF, which is done by connecting a CRF layer after UniLM [28]. When a predicted sequence has a high score, taking the output maximum probability value of each position corresponding to the tags to form a sequence is not necessarily the optimal result in line with the real situation, but also to consider the transfer probability between the tags. The CRF layer can add some rules to the last predicted labels to ensure that the predicted labels are legal, and at the same time, the transfer feature matrix in the CRF will take into account the sequentiality between the output labels, so as to ensure that the optimal sequence of output labels is obtained in compliance with the rules. In the process of model training, these rules and the transfer feature matrix can be learned automatically and do not need to be set by human.

In addition this task model is trained to set a separate learning rate for the CRF layer parameters, typically more than 100 times the learning rate of the pre-trained model, with an initial value set to 1×10^{-2}.

When training neural networks, often use a gradient descent based approach to continuously reduce the difference between the predicted value and the true value, and this difference is called the loss, the function to calculate this loss is called the loss function. Neural network models with different output types have different loss functions. For models with a CRF layer loss function CRF loss function consists of two parts, the score of the true path and the total score of all paths. The score of the true path should be the highest score of all the paths. The score of the path is obtained by adding the emission score and the transfer score. The specific formula is as follows:

$$P(Y \mid X) = \frac{\exp(score(X, Y))}{\sum_{\hat{Y}} \exp(score(X, \hat{Y}))} \tag{18}$$

$$score(X, Y) = \sum_{i=0}^{n} T_{y_i, y_{i+1}} + \sum_{i=1}^{n} E_{i, y_i} \tag{19}$$

$$Loss = -\log(P(Y \mid X)) \tag{20}$$

Where, X denotes the input sequence, Y denotes the true target sequence, $\hat{Y}$ denotes the sequence of possible output paths, E denotes the firing matrix scores, T denotes the transfer matrix scores, $score(X, Y)$ denotes the scores of each path, $P(Y \mid X)$ is the ratio of the true path

scores to the sum of the scores of all the paths, and Loss is the loss function of the CRF layer. The true path score should be the largest among all path scores, so maximize $P(Y \mid X)$, i.e., minimize Loss (which is inversely proportional to $P(Y \mid X)$), and train the model by backpropagation. For this task, the target sequence part needs to be masked.

The CRF layer uses the Viterbi algorithm to decode the sequence predicted by the model, and finds the optimal path through dynamic programming as the final output sequence of the model.

3.3. Japanese Grammar Error Correction Tasks

The Japanese grammar error correction task uses an end-to-end approach, where the sentence to be corrected is input into the model and the correct utterance is output directly, viewing the process as a translation problem, or as a sequence-to-sequence text generation problem. The task allows for the detection, localization, and correction of grammatical errors in a single pass.

The Japanese grammar error correction model implements an additional syntax error correction step over the error labeling task, i.e., syntax error detection, localization, and correction at the same time. The framework of the error correction model is an end-to-end seq2seq structure, where the Encoder side is the UniLM model, the Decoder side is a simple linear classification layer, and there is a Dropout layer between the Encoder side and the Decoder side.

The loss function used in the syntactic error correction model is the multicategorical cross-entropy loss function, calculated as follows:

$$Loss = -\sum_{i=1}^{K} y_i \log(p_i) \tag{21}$$

where K is the size of the dictionary, y is the uniquely hot encoding of the true label, y_i is 1 if the category is i and 0 otherwise, and p_i denotes the probability that the corresponding model output is of category i. In the syntactic error correction modeling technique loss, the part of the sentence sequence to be corrected is masked, i.e., the output loss of this part is not considered, and the loss comes only from the target sentence sequence part.

In the model decoding stage, cluster search is used, the beam size is set to 3, and the candidate result with the highest joint probability value is taken as the final output sequence of the model.

3.4. Multi-task learning

Multi-task learning is to train multiple related single tasks together to improve the accuracy of the model. Multi-task learning in this paper refers to the parameter optimization of backpropagation by adding the syntax error labeling task when training the syntax error correction model, viewing the UniLM as the Encoder end and sharing this part of the parameter, and summing the losses of the two tasks together.

The loss of multi-task learning model comes from different multiple sub-tasks, in this paper, we use the hyper-parameter task weight w to fuse the loss of syntax error labeling task and error correction task together, and the calculation formula is as follows:

$$Loss_{total} = w * Loss_1 + (1-w) * Loss_2 \tag{22}$$

where $Loss_{total}$ denotes the multi-task learning model loss, $Loss_1$ denotes the syntactic error labeling task loss, $Loss_2$ denotes the syntactic error correction task loss, and w is the manual setting of a hyperparameter between 0 and 1.

3.5. Error Correction System Architecture and Process

The realization of the function of the error correction system in this paper needs to rely on the support of the knowledge base and data, so in the design of the architecture of the error correction system in

this paper, it is necessary to establish a knowledge base of the Japanese language in the first place. Japanese language knowledge base and corpus are the basis for realizing natural language processing. The corpus is used to store language materials, explore language laws based on a large amount of language materials, and quantify them in computer language to realize language processing and error correction.

In this paper, in the design of error correction system architecture, we need to construct utterance rules to realize the analysis and error detection of utterances. Firstly, the binary syntactic synthesis rules are constructed, using the language knowledge base constructed above to classify and label each lexical category in the text, and according to the sequence of strings, the lexical element strings are obtained, which are specifically represented as:

$$Q = s_1|e_1, s_2|e_2, s_3|e_3, \cdots s_{i-1}|e_i, s_{i+1}|e_{i+1}, \cdots, s_n|e_n \tag{23}$$

where s_i is the lexical property, s_t before "|" is the smallest unit lexical element after the participle, and e_i after "|" is the most appropriate lexical property at the moment. In this paper, we interpret the syntactic form and organizational information of syntax through the principle of Japanese word collocation, and we combine the lexical properties of the above token set and linguistic rule structure, constituting a binary syntactic syntactic synthesis rule set. In Japanese syntactic structure collocation, in addition to the common binary syntax, there also exists ternary syntax, so it is also necessary to carry out the formulation of ternary syntactic syntax syntax synthesis rules and construct a ternary syntax syntax syntax syntax rule text library, which is similar to the binary syntax syntax syntax syntax rule set in its realization, but since the process of parsing the ternary syntax is a combination between the ternary words, in the general case of Japanese syntax, word collocation in Japanese syntax follows the principle of left and right lexemes being the same In general, the lexical matching in Japanese sentences follows the principle of left and right lexemes being the same, and takes the conjunction as the connecting link, and takes the modification relationship between Japanese sentences into full consideration. The scanning of Japanese sentence constituents is carried out in accordance with the rule of back-to-front matching, and after the lexemes are selected by scanning, the lexical matching is carried out with the newest lexemes, and the construction of the ternary vocabulary rules is completed with the combination of the language rules and the lexical token set.

In this paper, based on the rule perspective above, we analyze the syntax of natural language and design the text error recognition function, and derive the sentence decomposition rules through syntax and statistics. The modification between binary words is more complex compared with the modification of ternary vocabulary questioning, so when traversing the text of the linguistic material, we give priority to traversing the binary ruleset to get the parsing of the utterance. Since the recognition of misspelled words in Japanese requires full consideration of the contextual relationship and the expression intention of Japanese, words as the smallest independently operating linguistic unit do not have the identification of word boundaries, so the word normality in the sentence is judged by the degree of closeness of the connection between the current text and the context, and the sentence is given enough to complete the automatic word division, and the contextual features are utilized, and low-frequency hash-schema tuples are extracted by the hash-serial method, to check the possible error strings of Japanese words in the sentence, and to check the possible error strings of Japanese words. Japanese strings that may have errors, using mutual information to extract sentences with low degree of association in the sentence, and the formula for its degree of association is:

$$I(p_i, p_{i+1}) = \log_2 \frac{s(p_i, p_{i+1})}{s(p_i) \cdot s(p_{i+1})} \tag{24}$$

Where $I(p_i, p_{i+1})$ is the mutual information of the words used in the sentence, which indicates the degree of association of two neighboring words, p_i and p_{i+1} are the two neighboring words, the contextual relationship of the error is extracted to get the string suspected of having a misspelled word,

and the Japanese string is converted into a pinyin string to generate a similar pinyin candidate, so that the pinyin string can be decoded again to get the new Japanese string, and the result of the recognition of misspelled words is outputted.

Finally, the error correction model is constructed based on natural language processing to realize the text correction function of the error correction system. Compared with English text, since the entry of Japanese text is realized by encoding, the form of its non-standard expression is more complicated, which brings more working difficulty for the error correction system, in order to solve this problem, in order to correct Japanese text, it is necessary to convert Japanese into five-stroke encoding, encode the input and then decode it into Japanese, and then combine with contextual context and the binary model to carry out the correction of errors. After the text error recognition, the language knowledge base of this paper is used to deeply analyze the word collocation relationship in the text, and on the basis of this, the collocation of words is obtained, and according to the collocation of words, the error correction of homophones is carried out. After Japanese strings are converted to pinyin strings, automatic word splitting is performed on the pinyin stream to realize the pre-processing of pinyin string decoding, and in the process of decoding the candidate pinyin strings, the decoding is performed according to the given observation sequences, which are then passed through the matching with the training language materials. The expression of the best matching string obtained after specific decoding is:

$$w' = \arg\max(w \mid k, \lambda) \tag{25}$$

Where w' is the selected best matching string, k is the processed sequence of pinyin strings, and w is a possible Japanese string under the current pinyin string, taking into account the contextual contextual relationship as well as the frequency distribution in the training language material. In the model training stage, due to some limitations in the coverage of the training language material, the entire linguistic phenomenon cannot be completely described, resulting in data sparsity, and the smoothing method is utilized to assign all occurrences of strings a non-zero probability value to calculate the candidate string probability, which is given by Eq:

$$Y = \arg\max \sum_{i} s(t_i) s(t_i \mid t_{i-1}) \tag{26}$$

where t_i is the time, based on the application of this probabilistic model, the candidate string generated with the highest probability is combined with the original text to arrive at the correct sentence and complete the error correction.

In order to verify whether the error correction system architecture designed in this paper can realize the recognition and correction of textual errors in a more complete way and have a high recall and accuracy rate, after dividing and researching the textual error types, the errors in this paper, which accounted for 85% of the textual error types, were determined as the object of systematic error correction, and the specific design of the error correction system architecture is shown in Fig. 1. Before the error correction, the model, contextual context, Japanese collocation and other preparatory resources are first trained, and the binary rule set, ternary rule set, text knowledge base and error correction model constructed in this paper are deposited into the local files of the system, and the data preprocessing is carried out after inputting the text to realize the automatic word segmentation and lexical annotation, to judge the word-word continuity, lexical continuity, and then through the homophones and the contextual context, to carry out local homophones and contextual context for localized homophone error correction, and use Japanese collocation to discover remote errors to achieve word-level error correction.

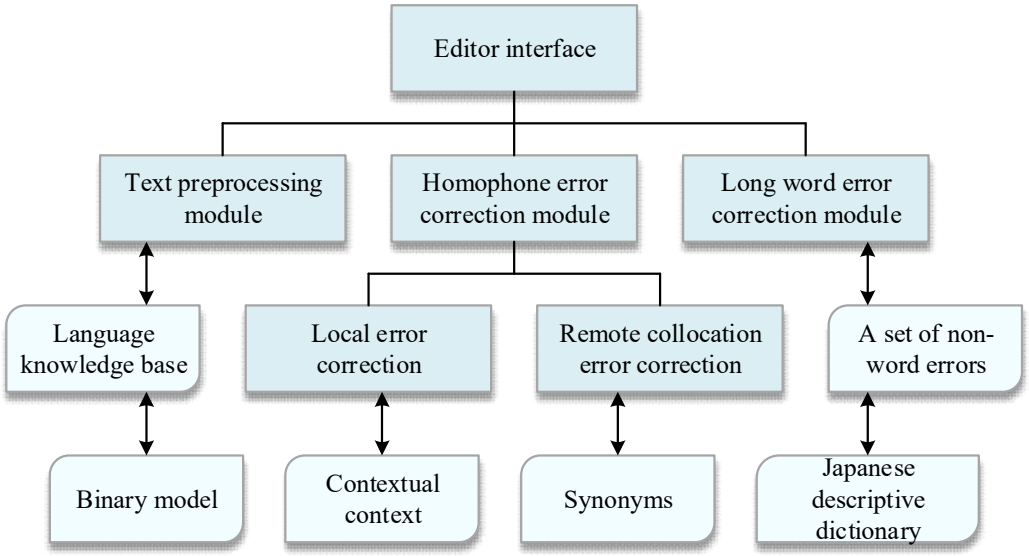

Fig. 1. Error correction system architecture

4. An innovative teaching model based on the Japanese grammar error correction model

Innovative teaching mode is the basic structure of all types of teaching activities established in practice by designing and organizing teaching under the guidance of certain teaching ideas or theories. Teachers are the executors of the teaching mode, and are the decisive force of teaching activities and teaching reform and innovation. Classroom teaching is the key channel for students to master knowledge, develop ability and expand quality. The traditional teaching classroom teaching and learning, the teacher sings a monodrama of the same type of thinking does not give full play to the student's subjectivity, did not follow the students' understanding of the laws of language, not the transmission of knowledge and ability to train the organic integration, so that the students lose interest in the course, and even produce an aversion to learning, and more inhibit the innovative thinking of students.

However, without students' active participation in classroom teaching, the quality of teaching cannot be improved. Therefore, to create a student-centered innovative teaching model must first innovate the concept of teaching, firmly establish a student-centered view of the main body of education; the development of students' abilities as the focus of the concept of quality of education, in order to improve the student's personality goal of education values, to achieve the student-centered teaching model of the pathway has a variety of ways to achieve the core of the students to actively participate in the teaching of the entire process of advocating the teacher-student cooperation in learning. Students are the active constructors and users of knowledge, and teachers are the guides, organizers, encouragers, advisers and evaluators of student learning. Therefore, it is necessary to cultivate students' awareness of independent learning and a new teaching mode of teacher-student interaction. In this paper, we design a system of Japanese grammar error correction model based on natural language processing technology, so as to establish a new teaching mode of teacher-student interaction.

Teachers and students should give full play to their own subjective initiative, enthusiasm and creativity as a prerequisite for teaching, create an equal, democratic and harmonious teaching environment as a condition, and take knowledge as a carrier, active participation, subjective co-development of a new type of teaching outlook or teaching mode. In the teaching method, the teacher's leading position and the students' subjective position should be guaranteed, and the students' independent mutual assistance must be guided and regulated by the teacher, therefore, the Japanese

grammar error correction model based on natural language processing technology can give the teacher a more scientific and accurate guidance to the students. It is a teaching that combines learning and creativity and promotes the growth of teaching and learning, which is very conducive to mobilizing students' initiative, enhancing students' creativity and promoting the development of students' innovative qualities.

5. Experimentation and analysis of Japanese language error correction models

5.1. Experimental analysis of data expansion

In the data expansion experiment section, the Lang-8, NUCLE, FCE, and CLEC corpora are expanded with data using the fusion rule and probabilistic approach. The training set of the learner corpus contains a total of about 2M parallel sentence pairs, and for the size of data expansion, we set the quantities {4M, 8M, 12M, 16M}. The original Lang-8, NUCLE, and FCE training sets are used to fine-tune the model of this paper, and when the performance of the model on the development set is no longer improved, it is then tested on the CoNLL-2014 test set, and the test results of the final model are shown in Table 1.

It can be found from the table. The performance of the model is rising as the size of the synthetic data becomes larger, and the main improvement to the model is reflected in the precision rate, which is 73.21% when the size of the data expansion is 16M, which is an improvement of 3.39%. At the same time, the recall rate also has some improvement, but it is not very obvious. In addition, when the scale of data expansion is increased to 16M, the performance improvement of the model is relatively reduced, which indicates that there is a bottleneck for the model to rely on the increase of synthetic data scale to improve the performance, and the data expansion strategy can be further optimized in the future so that the model can learn from richer error correction corpus and improve the precision and breadth of error correction.

Table 1. Data expansion experiment results

Training data	CoNLL-2014		
	P	R	F0.5
Learners' vocabulary	69.82	38.21	59.83
Learners' vocabulary + 4m synthetic data	71.79	38.39	61.16
Learners' vocabulary + 8m synthetic data	72.31	38.52	61.53
Learners' vocabulary + 12m synthetic data	72.9	39.08	62.14
Learners' vocabulary + 16m synthetic data	73.21	39.25	62.41

5.2. Analysis of model comparison experiments

In this subsection, the syntactic error correction models designed in this paper will be compared and analyzed with previous studies on CoNLL-2014 test set and JFLEG test set respectively.

For testing on CoNLL-2014 test set, four baseline models, denoted as Models 1~4, are selected in this paper as follows:

(1) Based on the Seq2Seq framework, using GRU as the main structure of the encoder-decoder and incorporating a nested attention mechanism at the word and character level.

(2) Based on the Seq2Seq framework, using multi-layer CNNs and attention mechanisms, while introducing a strategy of multi-model integration.

(3) A combination of SMT and NMT models is used, using n-gram language models in the SMT approach and Bi-GRU models in the NMT approach.

(4) A Transformer model with a copy mechanism was pre-trained with a denoised autoencoder.

The test performance of this model on the CoNLL-2014 test set is shown in Table 2, and after comparing with other baseline models, it can be seen that this model achieves leading results. The precision, recall and F0.5 values of this model are more than 10 percentage points higher compared to Model 1 as well as Model 2. The precision rate is 6.27 percentage points higher, the recall rate is 4.71 percentage points higher and the F0.5 value is 6.15 percentage points higher when compared to model 3. Compared to Model 4, which also uses the pre-training method, the model in this paper also achieves a small lead.

Table 2. The performance of this model in the conll-2014 test set

Model	Structure	P	R	F0.5
1	Nested-GRU	55.35	25.94	45.14
2	MLConv(4 ens.)+EO	62.39	27.55	49.78
3	SMT+Bi-GRU	66.85	34.51	56.21
4	Copy-Augmented Transformer+Pretrain	71.56	38.61	61.18
This model	UniLM+CRF	73.12	39.22	62.36

The test performance of this model on the JFLEG test set is shown in Table 3, and it can be seen that the GLEU score of this paper's model is slightly lower than that of Model 3, which incorporates the SMT approach, and its use of an n-gram language model improves the GLEU score to some extent, although it also imposes additional computational overhead. The model in this paper achieves a GLEU score that is nearly identical to that of Model 3 while not incorporating the additional SMT model, outperforming the other NMT models. The score of this model is also very close to the human score, with a difference of about one percentage point, proving that this model performs well in the metric

of error correction fluency.

Table 3. The performance of this model in the JFLEG test set

Model	Structure	GLEU
1	Nested-GRU	52.51
2	MLConv(4 ens.)+EO	57.89
3	SMT+Bi-GRU	61.59
4	Copy-Augmented Transformer+Pretrain	61.05
Human performance	—	62.68
This model	UniLM+CRF	61.52

5.3. Experimental Analysis of Students' Composition Critique

In order to further validate the performance of this paper's model on the grammar correction of actual Chinese students' Japanese compositions, 600 Japanese compositions of college students who are not majoring in Japanese were tested from the SET3 and SET4 libraries of the CLEC corpus, and the grammatical errors were classified according to the grammatical error categories. Table 4 shows the results of grammatical error correction of students' Japanese compositions by the model of this paper. In order to be more intuitive, the data in the table are presented using graphs, and the 3D bar chart for the classification of grammatical error correction is shown in Figure 2. It can be seen that the model in this paper has good performance on actual grammar correction of students' Japanese compositions. The average precision rate of grammatical error correction reaches 83.27%, the recall rate reaches 71.32%, and the F1 value reaches 76.85%.

Table 4. The error correction of the composition of the students

Error category	Error number	Correct number of errors	Correct error number	P	R	F1
ArtOrDet	32	30	29	0.8912	0.794	0.8509
Nn	59	61	55	0.9323	0.8731	0.9015
Npos	4	1	2	0.6662	0.5	0.5714
Pform	12	9	7	0.778	0.5385	0.6365
Pref	24	19	17	0.7617	0.6401	0.6954
Prep	23	26	17	0.8334	0.7305	0.7597
Rloc-	58	49	38	0.775	0.6907	0.73
Ssub	136	116	97	0.805	0.6878	0.7416
SVA	57	49	46	0.9386	0.8219	0.8768
Trans	6	2	-2	0.3331	0.2497	0.2857
V0	57	51	42	0.796	0.6841	0.7358
Vform	43	36	33	0.7936	0.6789	0.7202
Vm	25	17	13	0.8331	0.6006	0.6973
Vt	31	25	23	0.8082	0.6995	0.7499
Wci	7	1	2	0.4998	0.2002	0.2857
Wform	2	0	0	0.5002	0.2497	0.3333
WOadv	9	5	1	0.5995	0.2502	0.3526
WOinc	17	7	6	0.3641	0.2354	0.2856
Others	83	75	60	0.8355	0.7175	0.7719
All	685	579	486	0.8327	0.7132	0.7685

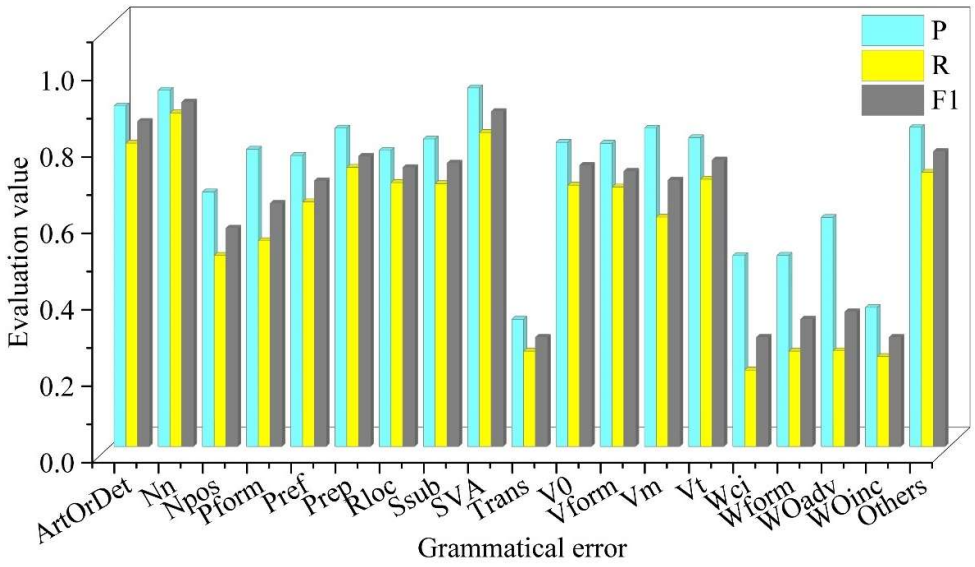

Fig. 2. A statistical chart of grammatical errors and errors

The accuracy rates of correcting the various types of grammatical errors in the figure are ranked from high to low and segmented, and the statistics are shown in Table 5. It can be seen that the model's accuracy rate exceeds 90% in correcting two kinds of errors, namely, noun singular-plural and subject-predicate consistency, and exceeds 60% for the vast majority of grammatical errors. It shows that the model structure designed in this paper takes into account both syntactic and semantic level information, while the use of hybrid attention mechanism makes the model can locate the keywords to be referred to more accurately when correcting errors, which leads to an increase in the accuracy rate of error correction.

Table 5. Correct precision segmentation of grammatical errors

Error correction accuracy	Grammatical error category
>90%	Nn, SVA
80%-90%	ArtOrDet, Prep, Ssub, Vm, Vt, Others
70%-80%	Pform, Pref, Rloc-, V0, Vform
60%-70%	Npos, WOadv
<60%	Trans, Wci, Wform, WOinc

The various types of grammatical error corrections were ranked from highest to lowest recall. The recall rate of correction of various types of grammatical errors is segmented, and the statistics are shown in Table 6. It can be seen that the model in the noun singular-plural and subject-verb consistency error recall more than 80%, and in the correction of some common grammatical errors also have a good recall performance, which is mainly due to the method of data expansion, the expanded data reflects the distribution of grammatical errors commonly committed by the students in the distribution of the categories of grammatical errors, which allows the model to improve the recognition of these errors, and thus obtain a good recall performance.

Table 6. Correct recall segmentation of grammatical errors

Error correction accuracy	Grammatical error category
>80%	Nn, SVA
70%-80%	ArtOrDet, Prep, Vt, Others

60%-70%	Pref, Rloc-, V0, Vform, Ssub, Vm
50%-60%	Npos, Pform
<50%	Trans, WOadv, Wci, Wform, WOinc

Another 1,000 Japanese compositions were further selected and manually grammar-corrected and scored by the team's Japanese language teachers, with the total grammar-correction score for a Japanese composition set at 15 points, and then the model's grammar-correction result scores were compared with the teachers' scores. The comparison results are shown in Figure 3. It can be seen that the average score of the teacher's grammar correction and the average score of the model's correction are 12.46 and 12.65 respectively, and the error between them is only 0.19 points, thus proving that the automatic error correction model of Japanese text grammar designed in this paper is of good practicability, and it can replace the teacher's grammar correction of Japanese compositions to a certain extent.

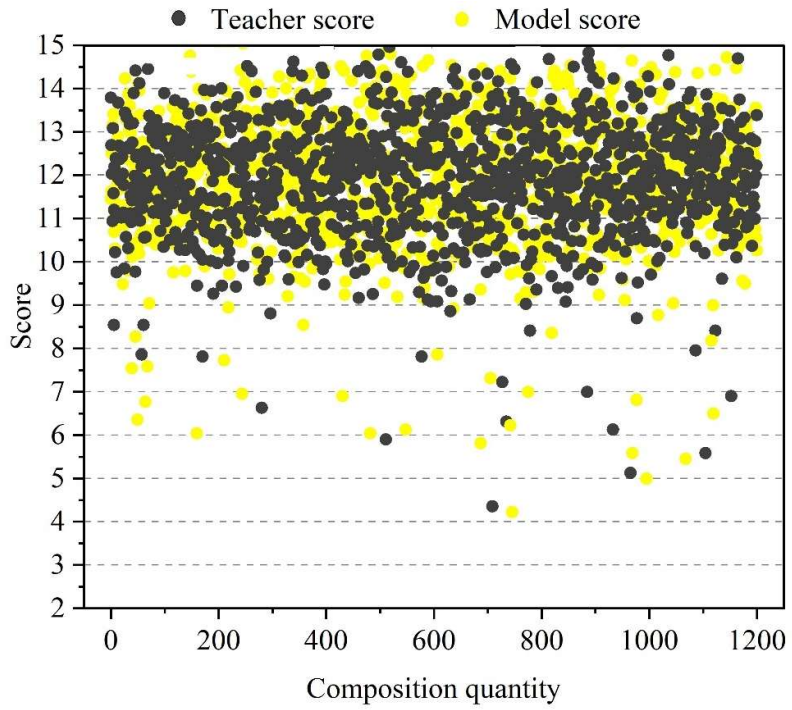

Fig. 3. Model score and teacher score contrast diagram

6. The effect of innovative teaching on the syntactic error rate of Japanese language learners

6.1. Pre- and post-test data analysis between groups

In order to further examine the experimental subjects' grammatical ability, the researcher will conduct a pre-test description of syntactic errors in the same Japanese writing materials for both classes. The results of the pre-test description are shown in Table 7, the frequency of syntactic errors between the control class and the experimental class are similar, all of them are the experimental class has higher frequency of syntactic errors than the control class, the total value of the experimental class has 7 more than the control class, and the percentage of the errors is 12.14% versus 15.18%, respectively.

Table 7. The statistical table of the pre-error measure

Syntax error	Class	Number	Error total	Error ratio/%

classification				
Phrase structure error	Cross-reference class	50	34	4.63
	Laboratory class	50	36	5.75
Clause error	Cross-reference class	50	22	3.03
	Laboratory class	50	23	3.67
Internal error	Cross-reference class	50	15	2.07
	Laboratory class	50	17	2.72
Sentence external error	Cross-reference class	50	17	2.34
	Laboratory class	50	19	3.04
Total value	Cross-reference class	50	88	12.14
	Laboratory class	50	95	15.18

Meanwhile, in order to further determine whether there is a difference in Japanese syntactic errors between the experimental class and the control class before the implementation of the writing experiment, the researcher conducted an independent samples t-test on the data of the pre-test of syntactic errors of the experimental class and the control class, and the results are shown in Table 8. In terms of phrase structure errors in the Japanese writing corpus of the experimental subjects, the Sig. (two-tailed) is 0.685, which is greater than 0.05, indicating that there is no significant difference between the two groups in the frequency of phrase structure errors. In terms of subordinate clause errors, Sig. (two-tailed) is 0.686, which is greater than 0.05, indicating that there is no significant difference between the two groups in the frequency of subordinate clause errors. In terms of inter-sentence errors, Sig. (two-tailed) is 0.524, which is greater than 0.05, indicating that there is no significant difference between the two groups in the frequency of inter-sentence errors. In terms of articulation errors, Sig. (two-tailed) is 0.946, which is greater than 0.05, indicating that there is no significant difference between the two groups in the frequency of articulation errors. In terms of total syntactic errors, Sig. (two-tailed) is 0.468, which is greater than 0.05, indicating that there is no significant difference between the two groups in terms of the frequency of total syntactic errors. In conclusion, there is no significant difference in syntactic errors at the grammatical level between the experimental class and the control class before the experiment, which can be used as subjects for this experiment.

Table 8. The independent sample t test table is tested before the syntax error

Syntax error	Levin variance equivalence test			T test						
		F	Sig.	t	df	Sig.2	D	Std.e	95% CI lower limit	95% CI upper limit
Phrase structure error	Assumed equal variance	0.094	0.755	0.412	88	0.685	0.075	0.175	-0.274	0.416
	Unassuming equal variance			0.404	80.125	0.685	0.074	0.174	-0.282	0.423
Clause error	Assumed equal variance	0.175	0.677	0.406	88	0.686	0.056	0.134	-0.215	0.325
	Unassuming equal variance			0.407	87.652	0.689	0.056	0.134	-0.215	0.324
Internal error	Assumed equal variance	1.175	0.278	0.645	88	0.524	0.067	0.106	-0.145	0.274
	Unassuming equal variance			0.645	86.269	0.525	0.057	0.105	-0.144	0.275
Sentence external error	Assumed equal variance	0.017	0.892	-0.065	88	0.946	-0.006	0.105	-0.208	0.195
	Unassuming equal variance			-0.065	86.254	0.945	0.006	0.104	-0.207	0.195
Total value	Assumed equal variance	1.065	0.306	-0.732	88	0.468	-0.235	0.326	-0.877	0.406
	Unassuming equal variance			0.728	85.465	0.465	-0.236	0.326	-0.875	0.411

In order to examine the effect of innovative teaching on syntactic errors among the experimental subjects' grammatical competence. The researcher issued a total of 100 post-test questionnaires, 100 valid questionnaires were returned, and the validity rate of the questionnaires was 100%. The post-test grammatical error rates in the relevant subjects' corpus were obtained. The posttest grammatical error rates are shown in Table 9. In the syntactic sub-dimension of grammatical competence, the experimental group made fewer errors than the control group. The total number of errors in the control group was 40 more than that in the experimental group, accounting for 11.76% and 7.46% of the total number of errors. In conclusion, after the innovative teaching experiment based on the Japanese error correction model, the number and percentage of syntactic errors in the experimental group were significantly smaller than those in the control group.

Table 9. The statistical description of the syntax error rate

Syntax error classification	Class	Number	Error total	Error ratio/%
Phrase structure error	Cross-reference class	50	32	4.32
	Laboratory class	50	18	2.86
Clause error	Cross-reference class	50	20	2.7
	Laboratory class	50	10	1.59
Internal error	Cross-reference class	50	17	2.33
	Laboratory class	50	8	1.27
Sentence external error	Cross-reference class	50	18	2.43
	Laboratory class	50	11	1.75
Total value	Cross-reference class	50	87	11.76
	Laboratory class	50	47	7.46

To further test whether there was a significant difference in syntactic errors in grammatical competence between the experimental and control groups in the post-experimental test, the researcher conducted an independent samples t-test for both, and the results are shown in Table 10. With regard to the phrase structure errors in the Japanese writing corpus of the experimental subjects in the post-test, the Sig. (two-tailed) is 0.025, 0.025, 0.043, and 0.045, which are less than 0.05, due to the unequal variance between the control class and the experimental class in terms of the frequency of phrase structure errors, subordinate clause errors, intersentence intra-sentence errors, and the total value of syntactic errors, respectively.

In terms of articulation errors, the Levene's test has a significance probability of 0.406, which is greater than 0.05, indicating that the variance between the control class and the experimental class in terms of the frequency of articulation errors is equal. Therefore, analyzing the data in the first row of the articulation error dimension, the Sig. (two-tailed) from the first row of data is 0.077, which is greater than 0.05, indicating that there is no significant difference between the two groups in terms of the frequency of articulation errors.

In summary, the innovative teaching model based on the Japanese grammar error correction model can effectively improve students' mastery of phrase structure, subordinate clauses, and internal articulation in syntax, but it fails to improve their mastery of articulation (external) in syntax.

Table 10. The independent sample t test table is tested after the syntax error

Syntax error	Levin variance equivalence test			T test						
		F	Sig.	t	df	Sig.2	D	Std.e	95% CI lower limit	95% CI upper limit
Phrase structure error	Assumed equal variance	5.094	0.025	-1.192	88	0.034	-0.165	0.135	-0.424	-0.105
	Unassuming equal variance			-1.227	85.225	0.025	-0.162	0.131	-0.421	-0.098
Clause error	Assumed equal variance	8.827	0.005	-1.526	88	0.02	-0.165	0.108	-0.385	-0.052
	Unassuming equal variance			-1.554	87.598	0.025	-0.165	0.108	-0.382	-0.046
Internal error	Assumed equal variance	8.142	0.004	-1.395	88	0.045	-1.335	0.098	-0.318	-0.055
	Unassuming equal variance			-1.408	87.996	0.043	-0.132	0.096	-0.315	-0.056
Sentence external error	Assumed equal variance	0.705	0.406	-0.415	88	0.077	-0.042	0.097	-0.236	0.152
	Unassuming equal variance			-0.416	86.598	0.077	-0.042	0.098	-0.235	0.153
Total value	Assumed equal variance	0.062	0.006	1.942	88	0.044	0.502	0.255	0.012	1.02
	Unassuming equal variance			1.955	87	0.045	0.501	0.256	0.009	1.02

6.2. Within-group pre and post-test data analysis

In order to verify whether there is a reduction in syntactic errors among the grammatical competence of the experimental group and the control group under the pre- and post-tests of the experiment, the researcher will analyze the within-group pre- and post-test data of the two groups.

As shown in Table 11, after the experimental class was taught with the innovative teaching model based on the Japanese grammar error correction model, the syntactic errors in the experimental class produced fewer errors, fewer percentages, and fewer total values in each dimension than in the pre-test. There were 48 fewer total values, a 7.72% decrease in the percentage.

Table 11. Syntax error rate of the experimental group

Syntax error classification	Cross-survey	Number	Error total	Error ratio/%
Phrase structure error	Premeasurement	50	36	5.75
	Posttest	50	18	2.86
Clause error	Premeasurement	50	23	3.67
	Posttest	50	10	1.59
Internal error	Premeasurement	50	17	2.72
	Posttest	50	8	1.27
Sentence external error	Premeasurement	50	19	3.04
	Posttest	50	11	1.75
Total value	Premeasurement	50	95	15.18
	Posttest	50	47	7.46

In order to further determine whether there is a significant difference in syntactic errors among the grammatical competence of the experimental group before and after the experiment, the researcher conducted a paired sample t-test on the data of the experimental group before and after the experiment using SPSS 21.0, and the results are shown in Table 12.

For phrase structure errors, subordinate clause errors, and intra-sentence errors, the Sig (two-tailed) is 0.001, 0.005, and 0.000, respectively, which is much less than 0.05, and for the total value of syntactic errors, the Sig (two-tailed) is 0.000, which is also less than 0.05. For articulation errors, the Sig (two-tailed) is 0.211, which is greater than 0.05. In summary, the innovative teaching model based on the Japanese grammar error correction model can effectively improve students' mastery of phrase structure, subordinate clauses, and sentence internalization in syntax. However, there is no effective improvement in the articulation part of syntax.

Table 12. Test sample t test before and after syntax error

Syntax error	Pair difference								
	Pre-survey and posttest	Mean	Std	Standard error of mean	95%CI lower limit	95%CI upper limit	t	df	Sig.2
Phrase structure error	Pre-survey and posttest	0.382	0.663	0.105	0.174	0.586	3.735	43	0.001
Clause error	Pre-survey and posttest	0.265	0.586	0.092	0.078	0.442	2.875	43	0.005
Internal error	Pre-survey and posttest	0.265	0.442	0.062	0.125	0.402	3.815	43	0.000
Sentence external error	Pre-survey and posttest	0.098	0.482	0.074	-0.052	0.256	1.272	43	0.211
Total value	Pre-survey and posttest	1.00	1.23	0.192	0.605	1.98	5.106	43	0.000

Table 13 shows the pre and post-test descriptive statistics of the syntactic error rate in the control group. The control group had two more errors in the inter-interior error post-test than in the pre-test, and the articulation error group had one more error in the post-test than in the pre-test, while the rest of the dimensions produced fewer errors, percentages, and total values than the pre-test, percentages, and total values in the pre-test as well. The total value decreased by 0.38%.

Table 13. Control group syntax error rate statistics

Syntax error classification	Cross-survey	Number	Error total	Error ratio/%
Phrase structure error	Premeasurement	50	34	4.63
	Posttest	50	32	4.32
Clause error	Premeasurement	50	22	3.03
	Posttest	50	20	2.7
Internal error	Premeasurement	50	15	2.07
	Posttest	50	17	2.33
Sentence external error	Premeasurement	50	17	2.34
	Posttest	50	18	2.43
Total value	Premeasurement	50	88	12.14
	Posttest	50	87	11.76

In order to further determine whether there is a significant difference in syntactic errors among the grammatical competence of the control group before and after the experiment in the conventional

mode, the researcher conducted a paired samples t-test on the data of the control group before and after the experiment using SPSS 21.0, and the results are shown in Table 14. In terms of the total value of phrase structure errors, subordinate clause errors, intra-sentence errors, articulation errors and syntactic errors, the Sig (two-tailed) is 0.059, 0.156, 0.085, 0.096, 0.503, respectively, which are not significant differences. In summary, the teaching under the conventional mode fails to effectively improve the mastery of the syntactic competence of the Japanese language.

Table 14. Test of the test of the cross-section syntactic error

Syntax error	Pair difference								
	Pre-survey and posttest	Mean	Std	Standard error of mean	95%CI lower limit	95%CI upper limit	t	df	Sig.2
Phrase structure error	Pre-survey and posttest	0.103	0.365	0.056	-0.003	0.209	1.946	43	0.059
Clause error	Pre-survey and posttest	0.043	0.203	0.028	-0.016	0.095	1.431	43	0.156
Internal error	Pre-survey and posttest	0.063	0.243	0.037	-0.007	0.132	1.765	43	0.085
Sentence external error	Pre-survey and posttest	0.053	0.136	0.034	-0.013	0.105	1.763	43	0.096
Total value	Pre-survey and posttest	0.265	0.568	0.084	-0.102	0.426	3.265	43	0.503

7. Conclusion

In this paper, based on natural language processing technology and deep learning technology, an effective and feasible automatic Japanese error correction model is designed to realize the innovation of Japanese language teaching mode in colleges and universities.

The average grammar error correction precision rate, recall rate, and F1 value of the model in this paper are 83.27%, 71.32%, and 76.85%, respectively, and it performs well in the task of correcting students' Japanese compositions.

Comparing the model's correction result score with the teacher's score, the error between the average score of the teacher's grammar correction and the average score of the model's correction is only 0.19 points, which proves that the automatic error correction model of Japanese text grammar designed in this paper has a good application feasibility, and it can reduce the teacher's pressure on grammar correction of Japanese compositions to a certain extent.

The experimental group has fewer errors than the control group in the grammatical error sub-

dimension in both the experimental group and the control group in the posttest of the intergroup experiment. In total, the control group had 4.3% fewer errors than the experimental group. The total value of syntactic errors in the experimental class in the within-group experiment was 48 fewer, a 7.72% decrease in the proportion, and the difference in the total value pre- and post-test datability was significant. The proportion of total value in the control group decreased by 0.38%, and there was no significant difference in total value pre- and post-test data. It shows that the innovative teaching mode based on the Japanese grammar error correction model can effectively improve students' mastery of Japanese syntactic ability, while the teaching in the conventional mode cannot effectively improve students' mastery of Japanese syntactic ability.

Funding

Shaanxi Province “The 14th Five-Year Plan” Education Science Planning Program of Shaanxi province (NO. SGH23Y2765).

References

[1] Cambria, E., & White, B. (2014). Jumping NLP curves: A review of natural language processing research. IEEE Computational intelligence magazine, 9(2), 48-57.

[2] Crowston, K., Allen, E. E., & Heckman, R. (2012). Using natural language processing technology for qualitative data analysis. International Journal of Social Research Methodology, 15(6), 523-543.

[3] Nadkarni, P. M., Ohno-Machado, L., & Chapman, W. W. (2011). Natural language processing: an introduction. Journal of the American Medical Informatics Association, 18(5), 544-551.

[4] Hirschberg, J., & Manning, C. D. (2015). Advances in natural language processing. Science, 349(6245), 261-266.

[5] Reshamwala, A., Mishra, D., & Pawar, P. (2013). Review on natural language processing. IRACST Engineering Science and Technology: An International Journal (ESTIJ), 3(1), 113-116.

[6] Yue, X., Di, G., Yu, Y., Wang, W., & Shi, H. (2012). Analysis of the combination of natural language processing and search engine technology. Procedia Engineering, 29, 1636-1639.

[7] Zhang, B., Yan, H., Wu, J., & Qu, P. (2024). Application of Semantic Analysis Technology in Natural Language Processing. Journal of Computer Technology and Applied Mathematics, 1(2), 27-34.

[8] Khurana, D., Koli, A., Khatter, K., & Singh, S. (2023). Natural language processing: state of the art, current trends and challenges. Multimedia tools and applications, 82(3), 3713-3744.

[9] Alhawiti, K. M. (2014). Natural language processing and its use in education. International Journal of Advanced Computer Science and Applications, 5(12).

[10] Gao, X. A., Liao, Y., & Li, Y. (2014). Empirical studies on foreign language learning and teaching in China (2008–2011): A review of selected research. Language Teaching, 47(1), 56-79.

[11] Teo, T., Hoi, C. K. W., Gao, X., & Lv, L. (2019). What motivates Chinese university students to learn Japanese? Understanding their motivation in terms of ‘posture’. The Modern Language Journal, 103(1), 327-342.

[12] He, L. (2024). Analysis of the Current Situation and Development Trends of Japanese Language Education in Chinese Universities. Transactions on Social Science, Education and Humanities Research, 4, 132-137.

[13] Demir, K., Czerniak, C. M., & Hart, L. C. (2013). Implementing Japanese Lesson Study in a Higher Education Context. Journal of College Science Teaching, 42(4).

[14] Gayevska, O., & Kravtsov, H. (2022). Approaches on the augmented reality application in Japanese language learning for future language teachers. Educational Technology Quarterly, 2022(2), 105-114.

[15] Hua, W. (2017, July). Study on the e-Learning Model Based on Japanese Teaching in Colleges and Universities. In 2017 7th International Conference on Social Network, Communication and Education (SNCE 2017) (pp. 914-918). Atlantis Press.

[16] Xu, X. (2020). An empirical study based on the teaching quality and related issues of Japanese education in colleges and universities. Journal of Contemporary Educational Research, 4(1).

[17] Shen, R. (2018). Japanese Literature Course Teaching Research Under College Education [C]. Institute of Management Science and Industrial Engineering. Proceedings of 2018 4th International Conference on Education & Training, Management and Humanities Science (ETMHS 2018). Clausius Scientific Press.

[18] Liu, C., & Chen, Q. (2021, February). Current situation and suggested measures of Japanese teaching in colleges and universities based on computer aid. In Journal of Physics: Conference Series (Vol. 1744, No. 3, p. 032051). IOP Publishing.

[19] Song, G. (2021). Sentiment analysis of Japanese text and vocabulary learning based on natural language processing and SVM. Journal of Ambient Intelligence and Humanized Computing, 1-12.

[20] Liu, J., Fang, Y., Yu, Z., & Wu, T. (2022, March). Design and construction of a knowledge database for learning Japanese grammar using natural language processing and machine learning techniques. In 2022 4th International Conference on Natural Language Processing (ICNLP) (pp. 371-375). IEEE.

[21] Wei, Y. (2023). The Development and Teaching Application of Japanese Peripheral Language Phenomenon Based on Big Data Corpus. ACM Transactions on Asian and Low-Resource Language Information Processing.

[22] Kasmaji, A. N. S., & Purwarianti, A. (2015, August). Employing natural language processing to analyse grammatical error in a simple Japanese sentence. In 2015 International Conference on Electrical Engineering and Informatics (ICEEI) (pp. 82-86). IEEE

[23] Jing, Y. (2020, November). Research on the application of artificial intelligence natural language processing technology in Japanese teaching. In Journal of Physics: Conference Series (Vol. 1682, No. 1, p. 012081). IOP Publishing.

[24] Guo, Z., & Jifeng, Z. (2021). Japanese translation teaching corpus based on bilingual non parallel data model. Journal of Intelligent & Fuzzy Systems, 40(2), 3731-3741.

[25] Waidyasooriya Hasitha Muthumala & Hariyama Masanori. (2024). Performance evaluation of Word2vec accelerators exploiting spatial and temporal parallelism on DDR/HBM-based FPGAs. The Journal of Supercomputing(12),17192-17211.

[26] Malik Muhammad Shahid Iqbal,Nawaz Aftab,Jamjoom Mona Mamdouh & Ignatov Dmitry I.. (2024). Effectiveness of ELMo embeddings, and semantic models in predicting review helpfulness. Intelligent Data Analysis(4),1045-1065.

[27] Yuanhang Zeng,Guangzhi Zhu & Xiao Zhu. (2025). Seq2Seq model with attention for predicting nonlinear propagation of ultrafast pulses in optical fibers. Optics and Laser Technology(PC),112014-112014.

[28] Mengyuan Ding,Chenjie Yang,Jianyi Liu,Xuguang Lan & Nanning Zheng. (2025). Relationship detection for manipulation in object stacking scene with fully connected CRF. Neurocomputing128661-128661.