International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 17 (2025) pp. 1-15
© Cerebration Science Publishing
https://doi.org/10.70917/ijcisim-2025-0274
Article

Optimization Strategies for Improving the Translation Level of English Translation Classrooms in Higher Vocational Colleges and Universities Supported by Data Analysis Technology Support

Yitian Zhang *

School of Foreign Languages, Zhengzhou Preschool Education College, Zhengzhou, Henan, 450000, China; zhangyitian@zzpec.edu.cn

Abstract: Higher vocational colleges and universities are more concerned about whether students can rapidly improve their skill level in actual teaching due to their employment-oriented characteristics. Data analysis technology has application advantages such as rapid response and intelligence, and the application of this technology to improve students' translation level can effectively meet the modernization reform needs of higher vocational colleges and universities. In this paper, students' English translation works are transcribed into the work archive as the research corpus. A computer intelligent detection and proofreading system based on the improved phrase translation model is built to detect and mark the errors in the students' translation work data. Complete the proofreading of translation error knowledge points with machine translation for students' reference. After 6 weeks of application, the students' English translation level in all 6 dimensions was improved to more than 80 points, which verified the optimization effect of using the system for the students' English translation level.

Keywords: data analysis; improved phrase translation; intelligent detection and proofreading; higher vocational colleges; translation level

1. Introduction

In the context of globalization, English translation has become an important tool to promote communication between different countries and cultures, and English translation education has also become the focus of the education field. However, with the diversified changes in society's demand for translation talents, the traditional English translation education model faces some challenges and urgently needs innovation and change [1]. Big data, or megadata, refers to large or complex data sets that are insufficiently processed by traditional data processing applications, and can also be defined as a large amount of unstructured or structured data from various sources [2-3]. Its data comes from a wide range of sources, including social media, sensor networks, and other various fields. Introducing big data in the English translation education model and mastering all kinds of information in the process of students' translation learning can provide accurate and scientific basis for teaching decision-making, thus enhancing the effectiveness of English translation talents [4-5]. The value of the big data-enabled English translation education model is mainly reflected in the following aspects:

Provide data support and enhance the pertinence of parenting. In conventional English translation teaching, it is difficult for teachers to understand the learning situation and personality characteristics of each student in a comprehensive and detailed way, so they adopt a more uniform teaching content and guidance methods, and it is difficult to accurately meet the diverse learning needs of students. A large amount of information generated by students in the process of English translation learning, such as their performance in classroom discussions, completion of homework after class, quality of translation works,

Copyright: © 2025 by the authors

and mastery of different types of translation topics, can clearly reflect the learning trajectory and knowledge weaknesses [6-7]. Based on the in-depth analysis of these data, teachers can more accurately grasp the learning style, learning progress and interest preferences of each student, so as to adopt more targeted teaching content and methods, and enhance the relevance of parenting [8-9]. For example, for students who are good at literary translation, rich in literary literacy and imagination, teachers can expand the learning content of literary genre translation skills to cultivate their higher level of literary appreciation and translation ability. This personalized teaching mode based on big data is conducive to truly realizing tailored teaching and helping each student grow and progress in English translation learning [10].

Deeply connect teachers and students and improve management accuracy. Various types of intelligent terminals can dynamically record all kinds of basic data in management work, deeply connecting teachers' teaching and management behaviors with students' learning behaviors and learning achievement status, so as to improve the accuracy of management [11-12]. First, relying on data collection technology, it can track the learning content, social needs, psychological conditions, personality preferences and other behavioral information left by students in the process of platform browsing, according to which it establishes the basic information database of students, providing an accurate source of data for management [13]. Secondly, based on data mining and intelligent analysis technology, teachers can extract relevant data that can fully reflect students' behavioral characteristics and ideological tendencies from the students' basic information database to provide support for the implementation of different education management services [14]. Thirdly, relying on the big data prediction function, teachers can quickly grasp the common characteristics and personality characteristics of the student population, providing a basis for the implementation of personalized teaching [15].

Innovative teaching mode to stimulate students' enthusiasm. The traditional English translation teaching mode has a low interaction rate, which stems from the lack of classroom vitality, and the teaching method is based on the lecture method. The English translation teaching mode empowered by big data is “student-centered”, and with the help of massive teaching resource library and intelligent platform, it can extend the classroom space and expand the interactive subject and interactive scope [16-17]. On this basis, the rational arrangement of teaching content, teaching methods, teaching evaluation, can make the layout of the classroom more three-dimensional, really let the students become the master of learning.

Expand the teaching space and improve the effectiveness of work. Teaching space refers to the scope involved in the development of teaching activities, including real physical space and network virtual space [18]. By creating a teaching ecology that integrates massive resources, personalized services, and data support, it can provide students with a more open, flexible, efficient, and expansive teaching space, thus extending the teaching space [19-20]. Under this educational field, students can independently choose learning contents and arrange learning plans, while teachers can efficiently and effectively integrate teaching resources and organize interactive activities, etc. [21]. Teachers can also accurately record the teaching process, effectively collect students' learning data, and form personalized evaluation and feedback to enhance the effectiveness of English translation teaching [22].

Literature [23] designed a public English automatic translation system and applied it to English translation teaching for the problems faced by traditional translation teaching in the era of big data, and deeply analyzed the feasibility and advantages of big data technology in English translation teaching. Literature [24] self-made online learning algorithm for English translation based on big data analysis technology, which reduces the complexity of big data in online English translation by introducing radial function, and achieves a better feedback effect in the practice of optimization of English translation teaching and provides an effective path for English translation teaching. In addition, literature [25] points out that the intelligent translation system based on big data and IoT technology has faster translation speed, higher accuracy and more rapid system response, and the intelligent translation system used in the study has a translation speed of 5.89 seconds when the number of sentences is 10,000, and the translation accuracy of the sentences reaches more than 95%. Literature [26] developed an English translation learning system by combining big data analysis and Internet of Things technology, taking data sensitivity and confidence as input variables of the system while integrating the trust value between users, analyzing the data in the English translation learning system, and the study made good progress and improved the learning effect of English translation teaching. Literature [27] analyzed the implementation effect of English translation flipped classroom under the big data model, for which a teaching experiment was designed and a questionnaire survey was combined to analyze the teaching effect in depth, and it was found that students with flipped classroom learning experience have improved their English grades, vocabulary, and speaking. In order to improve the quality of English translation education and the comprehensive ability of students, [28] tried to replace the traditional English translation teaching mode with a big data teaching platform, and the study pointed out that the online platform improved the quality of teaching through the integration of teaching data and resources, and the accurate analysis of student behavior realized better student management. Literature [29] designed a data-driven English translation teaching model, which gives full play to the teaching advantages of the model through the actual cases of university English translation teaching, and effectively avoids the many problems existing in traditional translation teaching. Literature [30] further explored the mechanism of big data affecting machine translation, summarized that big data affects machine translation from the aspects of language habits, translation of newborn words, language updating rules, special-purpose English, Chinese featured part-of-speech system, computer hardware and software, translation system and storage capacity, and machine translation teaching, and put forward effective suggestions for English translation teaching based on the results of the study.

From the above literature combing on English translation teaching, the previous research on the problems and countermeasures in traditional English translation teaching has achieved extensive research results, which boils down to one point: the application of big data technology can help students better master the translation skills and improve the translation level. In the face of the challenges and opportunities in the era of big data, English translation teaching needs to continue to explore and practice in order to cultivate more high-quality talents with excellent translation skills and cross-cultural literacy [31].

In this paper, on the premise of guiding the students to plan the transcription of English translation works and constructing an archive of English translation works, the data analysis technology is used to detect and proofread the corpus of these works. Translated sentences and phrases are detected one by one in terms of omission, spelling mistakes, numerical inconsistency, temporal inconsistency, etc., and the accuracy of detection is improved by relevant algorithms. The improved phrase translation model is introduced into the computerized intelligent system to carry out scientific machine translation of the target translated sentences. Afterwards, the students' translation errors are intelligently proofread with the machine translation, and the translation knowledge points that the students need to strengthen their mastery level are unearthed.

2. English Translation Level Improvement Method Based on Intelligent Data Analysis

2.1. Intelligent Detection and Correlation Analysis of Typical English Translation Errors

In actual translation teaching, different students make all kinds of errors in different translation assignments, and the types of errors are numerous and the number of errors varies greatly. In this section, several most typical errors are selected from the data collected by the English translation assignment submission system, and the types of errors include both traditional grammatical errors (e.g., coronary errors and spelling errors) and translation errors (e.g., omission errors) specific to students' translation review under the new translation teaching scenarios, and so on. Aiming at these errors, corresponding detection methods are realized respectively to better detect the errors in students' English translation assignments.

2.1.1. Missing Translation Error Detection Based on Information Content and Proximity Sets

Omission errors are also common in students' English translation assignments. Unlike grammatical errors such as coronal errors, omission errors are typical translation errors that cannot be recognized by common grammar checks. An omission error refers to the fact that part of the information in the original text before translation is missing from the translated sentence. Omitted translation will lead to incomplete information in the sentence, and if the relevant words and phrases are omitted in the key position, it may lead to a completely different meaning expressed in the end.

At present, there are fewer studies on the identification of omission errors, and more of them are analyzed from the point of view of the generation principle of machine translation based on neural networks, and there is no identification of omission errors for students in the translation reviewing scenario for the time being. Here, the omission error is not the error of machine translation of text, but the error made by students when they finish the review translation. Since the order of the machine translation is somewhat different from the actual order of the sentences, and also differs greatly from the order of the student's own sentence translation, it leads to omissions even if the student revises the translation against the machine translation. In the new scenario, the system not only collects students' review translation texts, but also contains the corresponding original text of the assignment and the machine translation text.

Additional auxiliary information can be provided for the identification of missed translation errors.

Omitted translations generally result in a change in the information content of the sentence, and therefore a measure of information content is introduced to measure the change before and after the sentence. For a given sentence with N words in the sentence, of which there are a total of T different words, the probability of occurrence and the amount of information in the sentence for a given word w_i is:

$$p_{w_i} = \frac{n_i}{N} \tag{1}$$

$$S = \sum_{i=1}^{T} \log_2 \frac{1}{p_{w_i}} \tag{2}$$

Considering the differences in language expression between Chinese and English, the same semantics but different expressions may occur after translation, and then it may not be accurate to compare the entropy change before and after translation from a single perspective only. Therefore, firstly, the entropy is directly calculated for the three groups of data: the original text, the machine translation and the students' assignments, and secondly, on the basis of the pre-processing in the previous section, the entropy is calculated again after additional processing of the three groups of data, including the unified symbol format, the removal of the sentence stop words (including the punctuation), the word splitting, the restoration of the word shape and the correction of spelling errors. The above six sets of data are saved for subsequent model processing.

Also considering the problem of same semantics but different expressions, certain methods are needed to construct the near-synonym set s of the sentence:

1) After the original text is subdivided into words, translate word by word to obtain the set a1 and add it into s;

2) Obtain the set t1 and add it into s after the machine translation of the participle words;

3) Obtain the near-synonyms of set a1 added into s;

4) obtaining the near-synonyms of set t1 added to s.

Some of the assignments in the original dataset also contain “essential words”, i.e., the teacher stipulates that the students must use the word when they review the translation, so whether or not the word occurs also affects the identification of missing translations. The occurrence of words was calculated by comparing the assignments with the set of near-synonyms. The sentence length of the original text of the assignment and the sentence length of the corresponding machine translation were also added as auxiliary information to measure word occurrences.

After the above features are processed, the xgboost model is used for the final discrimination, and finally the missed translation error monitoring model is constructed. Figure 1 shows the overall flow of the leakage error monitoring model construction.

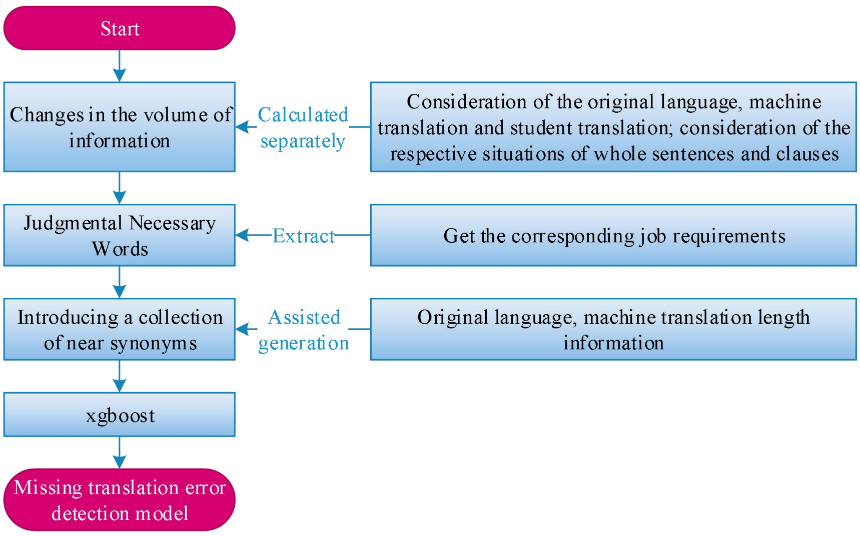

Figure 1. Construction Process of the Omission Error Detection Model.

2.1.2. Aspell-based improved spelling error detection

Spelling mistakes are one of the most common grammar-based errors, and they are extremely easy to make due to the presence of similarly pronounced words, words with similar spelling composition, or words of high complexity. Spelling mistakes are not only for people with weak English, but also for native speakers who have been using and writing English proficiently for many years.

To address this common error, there are many spell checking tools available, such as GNU Aspell, Ispell, Hunspell and Word Spell Checker. Among them, GNUAspell is a widely used open source spell checker. Compared with other spell checkers, Aspell not only finds spelling errors more accurately, but also provides additional suggestions for incorrect corrections to incorrect words.

However, Aspell still has room for improvement. For proper nouns such as names of people, places, organizations, etc., this checker is not able to recognize them well and often incorrectly labels them. And in combination with the above proposed specification for mislabeling, in order to label more reasonably, problems with the vocabulary of such proper nouns should be categorized as “proper noun errors or irregularities”, and therefore need to be processed and extracted separately for proper nouns. Meanwhile, although Aspell can make suggestions to improve errors, some of the suggestions have some deviation from the actual situation, and the suggestions are more likely to be affected by the editing distance, which leads to the change of the meaning of the words before and after the suggestions, and it is necessary to improve the reasonableness and accuracy of the suggestions. Figure 2 shows the step-by-step method of Aspell improvement.

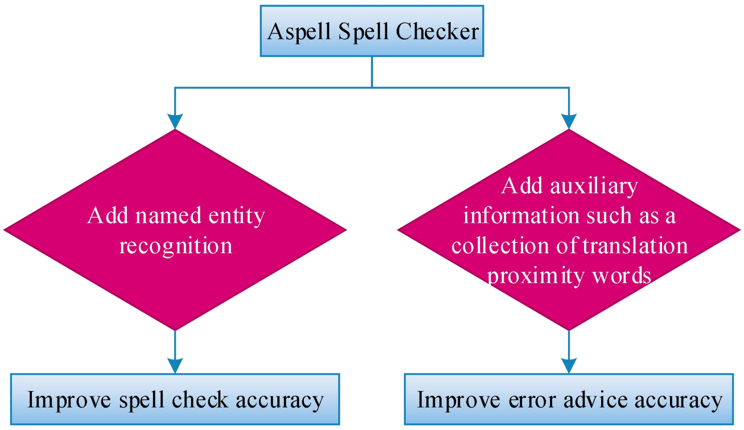

Figure 2. Improved Method of Aspell.

2.1.3. Digital Inconsistency Detection Algorithm

To assess whether the numerical information of the target text is consistent with the source text, the numerical information within the utterance needs to be extracted first. On the basis of sentence alignment, the numerical expressions of English-Chinese utterances are labeled, and in the process of numerical expression inconsistency detection, the numerical expressions are first extracted and added to the list. In general, the statements contain relatively little data information, so when comparing the consistency of numerical information, we can directly compare whether there are the same elements in the list of corresponding statements. Figure 3 shows the specific detection process of the numerical expression inconsistency detection algorithm.

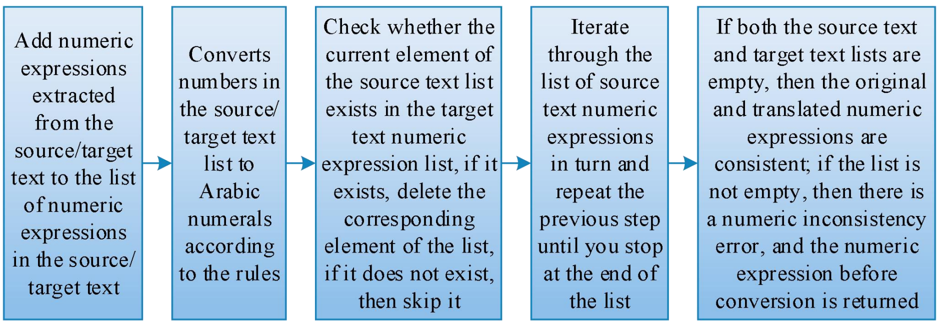

Figure 3. Process of Detecting Inconsistencies in Digital Expressions.

2.1.4. Temporal inconsistency detection algorithm

Time expression inconsistency detection means comparing the time in the source text and the target text utterances to audit the consistency of their descriptions. In the sentence alignment process, the main role of the time expression extracted from the English-Chinese statements is to assess the time expression consistency. Figure 4 shows the specific detection process of the temporal expression inconsistency detection algorithm.

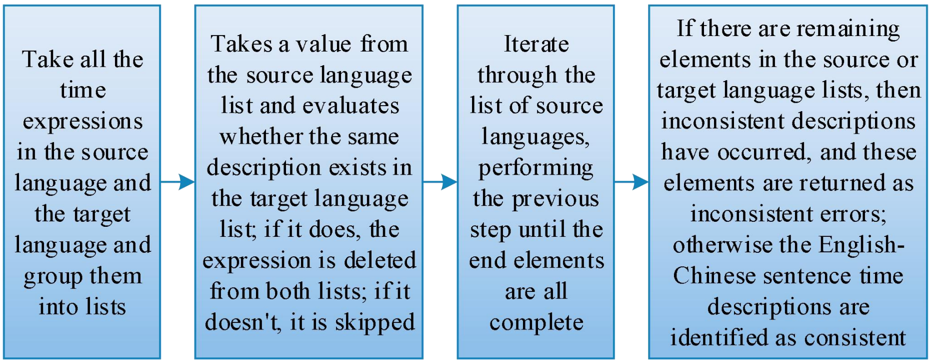

Figure 4. Process for Detecting Inconsistencies in Time Expression.

2.2. Intelligent English Translation Proofreading Based on Improved Phrase Translation Models

The conversion from a certain form of text to another form of text, i.e. the common point between English translation and translation proofreading. Therefore, computer intelligent proofreading is to translate towards the statement to be translated, compare and replace the conversion proofreading result with the initial translation result in order to realize computer intelligent proofreading. After inputting the phrases to be translated and students' translations, the computerized intelligent proofreading method based on the improved phrase translation model is utilized to translate the phrases and detect the errors of students' translations for one-to-one proofreading.

The computerized intelligent proofreading method based on the improved phrase translation model is:

$$G' = \arg\max_{e} W\left(\frac{G}{L}\right) = \arg\max_{e} W\left(\frac{L}{G}\right)\cdot W(G) \tag{3}$$

$W(G)$ mainly describes the vocabulary after proofreading.

While the accuracy of vocabulary translation in the translation result of machine translation method needs to be further improved, the computer intelligent translation method pays much attention to the accuracy of vocabulary translation, i.e., $W(G)$ accuracy. Therefore, based on the optimization of the above formula to complete the computer intelligent proofreading, namely:

$$G' = \arg\max_{e} W\left(\frac{L}{G}\right)^{\sigma}\cdot W(G)^{\lambda} \tag{4}$$

σ denotes $W\left(\frac{L}{G}\right)$ weight weights; λ denotes $W(G)$ weight weights.

In order to facilitate the elaboration of the computerized intelligent proofreading method, L denotes the vocabulary to be translated and proofread, and G denotes the vocabulary after proofreading. It is set that the to-be-translated proofreading vocabulary contains n characters characterized by L_1^n , the characters are relative to the vocabulary of the improved phrase translation model, and the post-proofreading vocabulary contains m characters characterized by G_1^m . Divide L_m^n into random c strings, characterized by L_1^c, with strings relative to the phrases of the improved phrase translation model; divide the proofread vocabulary into random c strings, characterized by G_1^c . Based on this, we obtain the extended formula of the above equation, i.e:

$$
\begin{aligned}
G' &= \arg\max_{e} W\left(\frac{G}{L}\right) \\
&= \arg\max_{G \cdot W_1'^c} \sum_{W_1'^c} W\left(\frac{L_1^c}{L_1^n}\right) \cdot W\left(\frac{G_1^c}{L_1^c, L_1^n}\right) \\
&\Rightarrow \arg\max_{G \cdot W_1'^c} \sum_{W_1'^c} W\left(\frac{L_1^c}{L_1^n}\right) \cdot W\left(\frac{L_1^c}{G_1^c}\right)^{\sigma} \cdot W\left(G_1^m\right)^2
\end{aligned} \tag{5}
$$

In computerized intelligent proofreading, the focus is on seeking the best method of dividing the vocabulary to be translated and proofreading in detail towards the results of the division, in order to obtain the proofreading results arranged in an orderly manner, which is organically combined with the method of elaboration of the above formula, in order to seek for vocabulary corresponding to the vocabulary to be translated and proofread, so as to realize the computerized intelligent proofreading.

2.3. Editing the archive of students' English translation works

2.3.1. Corpus editing and retrieval tools

The retrieval tool for the archive of students' translation works is Antconc software; the retrieval tool for the English-Chinese parallel corpus is Paraconc software; the text organizer used for text sorting is Fenglin Text Organizer; and the tool used for automatic lexical segmentation is Tokenizer.

In order to meet the analysis of grammatical features in the teaching of English translation, it is sometimes necessary to lexically assign codes to the corpus, and the software chosen is Treetagger, which, according to the research report of the researchers and developers, has an accuracy rate of more than 95% in assigning codes to the English language.

2.3.2. Transcription of the corpus

The corpus source of the archive of students' English translation works is the trial translation works of students in English translation teaching activities. In order to ensure the authenticity of the corpus, students are required to carry out time-limited translation and upload the data centrally after the trial translation.

In the practice of English translation teaching, due to the difference in the number of students, the corpus transcription can take different methods. If the number of students is small, the transcription can be done by teachers or teaching assistant students; if the number is large, it is recommended that students do it by themselves. However, before the corpus is formally transcribed, students need to be trained in corpus organization.

The basic requirements for transcription are:

1) When transcribing, record the following basic information in the text header: genre type, grade, year of enrollment, type of task, gender, self-diagnosis level, date of collection, and sentence order;

2) No spaces before each paragraph, only one space between words, and one space after punctuation;

3) Precede each sentence with a serial number, e.g., if the first sentence is a Chinese-to-English translation, it is labeled <CE1>, and so on. There can be no space between the CE or EC and the number.

2.4. Teaching process and steps

The teaching process of English translation classroom in higher vocational colleges proposed in this paper is based on the constructed archive of students' English translation works, and the scaffolding teaching mode is proposed by utilizing the constructivist translation teaching principle.

The teaching process is divided into three parts: before class, during class and after class. Before class, the teacher builds scaffolding, centers on the overall goal of English translation course design, builds a knowledge structure corresponding to translation tasks, selects translation tasks in the existing archive of English translation works, and makes a good classroom design. Students carry out preliminary translation according to the tasks issued by the teacher, and summarize and record the problems in the process of trial translation. The in-class session is divided into 3 main sections, 1) the teacher explains the relevant translation skills and the precautions for corpus transcription, and assists the students in transcribing the translations into the archive. 2) the teacher uses the intelligent detection system to intelligently detect translation errors in the students' translations, and leaves it to the students to complete the discussion of the corresponding errors and the second translation modification. 3) the teacher, through the computerized intelligent proofreading method based on improved phrase translation model, accurately proofreads the translations. Through the computerized intelligent proofreading method based on the improved phrase translation model, the teacher can accurately proofread the differences between the students' secondary translation works and the computerized translation, and guide the students to find the best English translation method. At the end of the class, based on the results of the intelligent testing and proofreading and the corpus of the English translation archive, students complete the English translation exercises after the class on the basis of acquiring translation skills and strategies, and mark them well for filing, so that they can create their own translation archive at the end of the semester.

3. The Practice of Enhancing English Translation Level Supported by Data Analysis Technology

3.1. Verification of Computerized Intelligent Proofreading Effect Based on Improved Phrase Translation Models

3.1.1. Verification of the Detection Effect of Improved Phrase Translation Models

In order to verify the performance of the designed improved phrase translation model, the experiment labels the corpus data from the students' English translation work archive, constructs the dataset, and applies it to the improved phrase translation model for training. Figure 5 visualizes the final model training results. With the increasing number of iterations, the loss function of the improved phrase translation model keeps decreasing from 6.417 to 0.520, and the accuracy keeps increasing from 66.968% to 96.465%. This shows that the detection accuracy of the model can be improved after training the model.

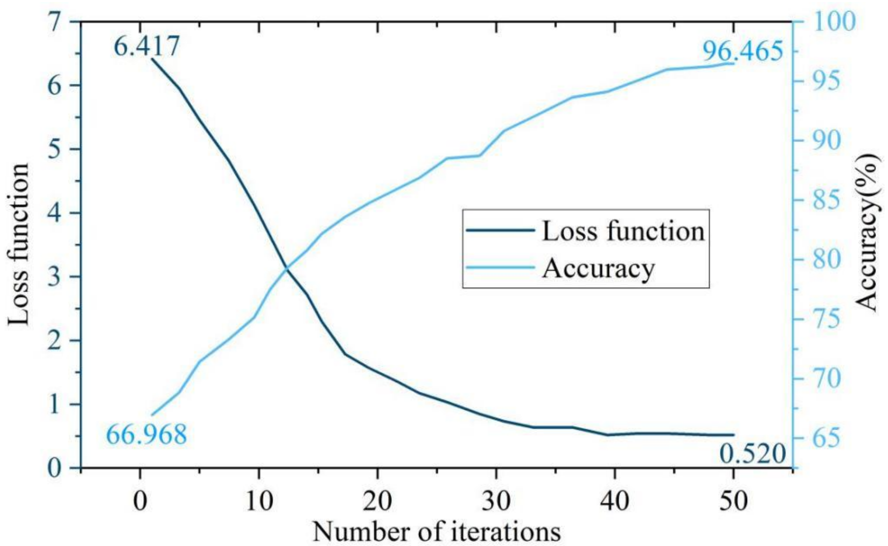

Figure 5. Improve the training results of the phrase translation model.

Table 1 shows the results of the comparison between the improved phrase translation model and the traditional detection model. The results of the improved phrase translation model are 0.887, 0.851, 0.865, 0.842 in the four evaluation indexes, namely, similarity, precision, recall, and $F_{0.5}$, which are higher than the other four comparison models. Using the trained improved phrase translation model, the students' translation corpus data can be fully utilized, from which different semantic and syntactic information can be obtained and compared with the computer's translation results, so as to realize the accurate detection and proofreading of English translation errors.

Table 1. Comparison of this model with traditional detection models.

Model	Indicators			
	BLEU	P	R	$F_{0.5}$
GEC	0.485	0.493	0.490	0.446
GED	0.502	0.527	0.515	0.538

MLM	0.539	0.543	0.579	0.594
GEC	0.611	0.670	0.623	0.658
Improve the phrase translation model	0.887	0.851	0.865	0.842

3.1.2. Verification of the effectiveness of intelligent computerized proofreading

The improved phrase translation model is deployed into the computer system to realize the intelligent detection and proofreading of students' English translation works. Still taking the dataset of 3.1.1 as an example, Table 2 shows the results of intelligent computerized detection and proofreading based on the improved phrase translation model. Whether it is omitted translation, spelling mistakes, or inconsistent translation of numbers and time, the scores of precision, recall and $F_{0.5}$ of intelligent detection and proofreading using the computer based on the improved phrase translation model exceed 0.800, and all the scores are in the range of 0.800-0.900, which makes the effect of error detection and proofreading more stable.

Table 2. Computer intelligent detection and proofreading results.

	Translation error	Testing	Translation error	Proofreading
Type of Error	P	$F_{0.5}$	P	$F_{0.5}$
	R		R	
Omission of translation	0.854	0.892	0.832	0.835
	0.813		0.815	
Spelling mistake	0.839	0.847	0.893	0.862
	0.827		0.891	
Figures inconsistency	0.895	0.856	0.835	0.871
	0.841		0.847	
Time inconsistency	0.860	0.890	0.826	0.883
	0.855		0.850	

3.2. Practice of Improving Translation Level in English Classroom

3.2.1. Preparation of materials for English translation

After verifying the effectiveness of the computerized intelligent detection and proofreading system based on the improved phrase translation model, the practice of improving translation level in English classroom was set up. The translation level test was conducted both before and after the practice to measure the students' translation level by means of written translation. The practice was conducted among the freshman (1) class of English translation majors in I higher vocational college. A total of 15 students participated in the 6-week translation practice. Teachers incorporated the translation knowledge points that students were prone to errors into the independent study materials and distributed them to students on a weekly basis for practice. Table 3 shows the statistical results of the translation knowledge points incorporated in the independent study materials. A total of 15 error-prone English translation knowledge points were integrated into the independent study materials, and the number of times each knowledge point was integrated in weeks 1-6 ranged from 19 to 30 times, which comprehensively examined students' English translation level.

Table 3. Statistics of translation knowledge points incorporated in the materials.

Knowledge Point	Week						Total
	1	2	3	4	5	6	
Co-translation(a1)	1	1	2	1	1	1	7
Increase translation(a2)	3	1	1	2	1	1	9
Proper translation(a3)	2	1	1	1	2	1	8
Split translation(a4)	1	1	1	1	2	2	8
Avoid omissions in translation(a5)	1	2	1	2	2	1	9
Be true to the original text(a6)	3	1	1	1	2	4	12
Part-of-speech conversion(a7)	1	1	1	1	2	2	8
Correct tense(a8)	1	1	3	1	4	1	11
Imperative mood(a9)	1	1	1	1	1	1	6
Matched accurately(a10)	2	1	2	1	1	2	9
Accurate use of articles(a11)	3	1	1	1	1	2	9
Vocabulary diversity(a12)	2	2	1	1	1	1	8
Spelling accurated(a13)	4	2	1	3	1	3	14
Consistent numbers(a14)	3	2	2	2	3	2	14
Reorder the sentence structure(a15)	2	1	2	2	1	3	11
Total	30	19	21	21	25	27	143

3.2.2. Changes and analysis of data from teaching experiments

Based on the students' weekly performance in English translation assignments, the researcher has made detailed statistics on the students' scores on the independent practice materials for 6 weeks. In this part, the changes in the students' translation scores and the changes in their knowledge acquisition will be explained. Table 4 shows the statistical results of the 15 students' translation scores for 6 weeks.The translation practice scores of the 15 students showed an increasing trend of change during the 6 weeks of practice. In weeks 1-4, the students' translation practice scores always remained below 80 although they kept growing, and it was speculated that the students might have experienced a short bottleneck period when using the computerized intelligent detection and proofreading system based on the improved phrase translation model. It was not until weeks 5 and 6 that the students' scores on the translation materials began to show an increase of about 5 or 6 points, and finally in the translation materials of week 6, the scores of all 15 students exceeded 90 points, indicating that the system's data analysis technology has a certain effect on the improvement of the students' translation level.

Table 4. Statistics of translation scores of 15 students over a period of 6 weeks.

Student ID	Week					
	1	2	3	4	5	6
1	71.046	73.995	77.441	79.382	84.629	90.723
2	70.293	73.241	76.687	78.634	83.942	90.036
3	71.935	74.883	78.328	80.275	85.583	91.677
4	71.374	74.322	77.768	79.715	85.023	91.117
5	70.743	73.691	77.131	79.078	84.386	90.485
6	70.182	73.713	77.159	79.106	84.414	90.508
7	71.232	74.108	77.554	79.501	84.809	90.903

8	72.394	75.342	78.788	80.735	86.043	92.137
9	73.194	76.142	79.586	81.533	86.841	92.935
10	73.281	76.229	79.675	81.622	86.937	93.031
11	71.392	74.314	77.765	79.712	85.024	91.118
12	70.188	73.136	76.582	78.529	83.837	90.931
13	71.283	74.231	77.647	79.594	84.902	90.996
14	71.485	74.437	77.883	79.783	85.091	91.185
15	71.325	74.273	77.719	79.676	84.984	91.072

The system's detection and proofreading labels for the students' incorrect knowledge points are accessed to visualize the trend of the students' translation knowledge point mastery. Figure 6 summarizes the students' knowledge point errors for 6 weeks. Similar to the scoring pattern, the number of students' knowledge point errors shows a decreasing trend from week 1 to week 6, and by week 6, basically all the knowledge point errors are around 0 times, and only a small number of knowledge point errors are 1-4 times, but the overall error rate has reached a very low level. It shows that the use of the system's data analysis technology can assist students to see the specific translation errors in their own translation materials and make targeted proofreading modifications, thus improving their mastery of the knowledge point.

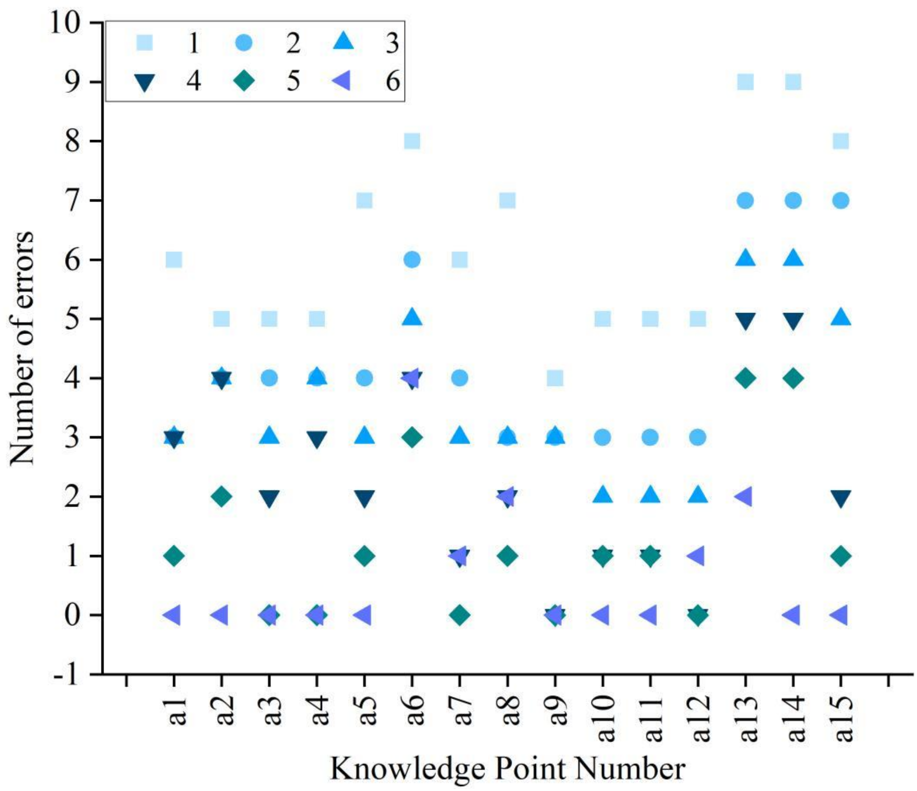

Figure 6. Students' mistakes in knowledge points over a period of 6 weeks.

3.2.3. Before and after comparison of students' English translation level

The students' English translation levels before and after the practice were compared to determine the students' improvement in translation level from the six evaluation dimensions. Because of the translation form of written examination, the students' final examination scores were used to represent their English translation level in each dimension. Table 5 shows the comparison of the 15 students' English translation level scores before and after the practice. Before the practice, the scores of the six dimensions of the translated works of the 15 students were no more than 80, while after the practice, the scores of the same six dimensions increased to more than 80, and the highest even reached 92.723. The scores show that the students' translation level has been greatly improved with the support of the system.

Table 5. Score of English translation proficiency before and after practice.

Student ID	Evaluation dimension					
	Language Structure Comparison and Transformation	Text production	Function realization and aesthetic expression	Translation Strategy	Personalized traits	Learning style in translation studies
Before						
1	70.394	69.346	71.399	71.017	68.539	71.631
2	70.273	69.225	71.278	70.896	68.411	71.503
3	70.291	69.243	71.296	70.914	68.429	71.521
4	71.385	70.337	72.439	72.057	69.572	72.664
5	71.094	70.046	72.099	71.717	69.232	72.324
6	71.446	70.398	72.451	72.069	69.584	72.676
7	70.198	69.415	71.468	71.086	68.601	71.693
8	71.345	70.297	72.365	71.983	69.498	72.529
9	71.293	70.245	72.298	71.916	69.431	72.523
10	70.936	69.888	71.941	71.559	69.004	72.096
11	71.245	70.197	72.825	72.443	69.958	73.075
12	71.237	70.189	72.242	71.086	68.601	71.693
13	70.456	69.408	71.461	71.079	68.594	71.686
14	71.99	70.942	72.995	72.613	70.128	73.292
15	71.236	70.185	72.243	71.852	69.371	72.464
After						
1	84.302	86.351	85.524	89.019	88.725	90.574
2	81.449	83.498	82.671	86.166	85.872	87.721
3	83.294	85.343	84.516	88.011	87.717	89.566
4	84.371	86.042	85.215	88.751	88.457	90.306
5	83.305	85.354	84.527	88.022	87.728	89.577
6	84.298	86.347	85.152	88.647	88.353	90.202
7	83.107	85.156	84.329	87.824	87.553	89.401
8	83.495	85.544	84.717	88.212	87.918	89.767
9	84.219	86.268	85.441	88.936	88.642	90.491
10	85.301	87.335	86.508	89.003	88.709	90.556
11	89.274	88.323	87.496	89.991	89.697	91.547
12	84.395	86.444	85.617	89.112	88.818	90.661

13	85.391	87.464	86.637	90.132	89.838	91.688
14	89.540	89.589	88.762	92.257	91.963	93.812
15	86.381	88.473	87.641	91.136	90.872	92.723

3.3. The Role of Computer Intelligent Application on Translation Level Enhancement

The computerized intelligent detection and proofreading system based on the improved phrase translation model affects students' English translation level in 4 dimensions: error detection, translation proofreading, repetition practice, and error marking. These 4 dimensions are used as independent variables, and students' translation scores are used as dependent variables to conduct multiple regression analysis to study the degree of influence of each dimension on students' translation level. Table 6 shows the regression analysis of the 4 dimensions of the system on students' translation achievement. The regression results show that the tolerances are all greater than 0.7 and the variance inflation factor is less than 2, indicating that there is no covariance problem in the 4 dimensions. $R^2 = 0.712$, which means that the regression model is well fitted and explains 71.20% of the variance in the translation scores. The most significant effect on students' English translation level is translation proofreading (β=0.311,p=0.001), followed by error detection (β=0.302,p=0.001). It indicates that the more powerful the system's translation proofreading and error detection functions are, the greater the likelihood that students' translation scores will improve.

Table 6. Regression analysis of 4 dimensions on students' translation scores.

Regression model		Fitting index			Coefficient			Collinearity diagnosis	
Independent variable	Dependent variable	R	R^2	F	β	T	P	Tolerance	Variance inflation factor
Constant	-				-	7.039	0.001	-	-
Error detection	Translation test scores	0.685	0.712	73.401 ***	0.302	7.503	0.001	0.794	1.046
Translation proofreading		-			0.311	5.692	0.001	0.832	1.532
Repetition practice					0.229	5.404	0.001	0.871	1.479
Error marking					0.158	5.193	0.001	0.746	1.538

Note: *** indicates p < 0.001

4. Conclusion

This paper introduces the improved phrase translation model, constructs a computerized intelligent detection and proofreading system, and analyzes the data of students' translation works to find the direction of optimization. Among the four influences of the system on students' translation performance, translation proofreading and error detection have the greatest influence, with β reaching 0.311 and 0.302 respectively.When teachers apply the system to assist students in higher vocational colleges to improve their English translation level, they should focus on the students' incorrect knowledge points and inaccurate translations, and concentrate on choosing suitable independent practice materials so that students can practice targeted a particular translation strategy or translation knowledge point.

References

1. Al-Hadithy, T. M. (2015). The traditional vs. the modern translation classroom: A need for new directions in the UAE undergraduate translation programs. Procedia-Social and Behavioral Sciences, 192, 180-187.

2. Tuken, A., Abbas, Y. M., & Siddiqui, N. A. (2025). Utilizing big data and categorical boosting modeling methodology to interpret the load-deflection behavior of steel fiber-reinforced concrete beams. Engineering Applications of Artificial Intelligence, 148, 110377.
3. Mayer-Schönberger, V., & Ingelsson, E. (2018). Big Data and medicine: a big deal?. Journal of internal medicine, 283(5), 418-429.
4. Jie, L. (2021, August). Modeling and analysis of translation ability of applied English talents based on data mining algorithm. In 2021 International Conference of Social Computing and Digital Economy (ICSCDE) (pp. 211-214). IEEE.
5. Waheed, H., Hassan, S. U., Aljohani, N. R., Hardman, J., & Nawaz, R. (2020). Predicting academic performance of students from VLE big data using deep learning models. Computers in Human behavior, 104, 106189-106189.
6. Schulze, M., & Scholz, K. (2018). Learning trajectories and the role of online courses in a language program. Computer assisted language learning, 31(3), 185-205.
7. Rizvi, S., Rienties, B., & Rogaten, J. (2018, October). Temporal dynamics of MOOC learning trajectories. In Proceedings of the first international conference on data science, e-learning and information systems (pp. 1-6).
8. Ning, J., Ma, Z., Yao, J., Wang, Q., & Zhang, B. (2025). Personalized learning supported by learning analytics: a systematic review of functions, pathways, and educational outcomes. Interactive Learning Environments, 1-23.
9. Yuan[1], L., Hou[1], B., Feng, D., & Rahman[1], S. S. A. (2025, June). Predictive Analytics in Education: Leveraging Al for Personalizing Student Learning Trajectories. In Proceedings of the 2nd International Conference on Educational Development and Social Sciences (EDSS 2025) (Vol. 924, p. 309). Springer Nature.
10. Li, R., Kirliauskienė, R., Sun, Y., Dong, S., & Zhang, L. (2022). Psychological quality of piano players based on big data algorithm. Wireless Communications and Mobile Computing, 2022(1), 7237099.
11. He, H., Yi, S., & Liu, W. (2020). Intelligent learning model of financial spoken English teaching based on BPTT algorithm and LSTM network model. Journal of Intelligent & Fuzzy Systems, 39(4), 4835-4846.
12. Escolano-Perez, E., & Losada, J. L. (2024). Using artificial intelligence in education: decision tree learning results in secondary school students based on cold and hot executive functions. Humanities and Social Sciences Communications, 11(1), 1-13.
13. Gasser, P., Grajeda, A., Cordova, J. P., La Fuente, I., Cordova, P., Naranjo, H., & Sanjinés, A. (2025). Mental cost in higher education: A comparative study on academic stress as a predictor of mental health in university students during and after the COVID-19 pandemic. Cogent Education, 12(1), 2445968.
14. Jiang, W. (2021, September). Research on the construction and application of college ideology network system based on data mining. In 2021 4th International Conference on Information Systems and Computer Aided Education (pp. 2430-2434).
15. Shou, Z., Li, Y., Li, D., Mo, J., & Zhang, H. (2025). Classroom network structure learning engagement and parallel temporal attention LSTM based knowledge tracing. PloS one, 20(4), e0320303.
16. Hmoud, M., Daher, W., & Ayyoub, A. (2025, July). From experience to engagement: a mixed methods exploration of learning environments using artificial intelligence and extended reality. In Frontiers in Education (Vol. 10, p. 1617132). Frontiers.
17. Cui, Y., Ma, Z., Wang, L., Yang, A., Liu, Q., Kong, S., & Wang, H. (2023). A survey on big data-enabled innovative online education systems during the COVID-19 pandemic. Journal of Innovation & Knowledge, 8(1), 100295.
18. Byers, T., Imms, W., & Hartnell-Young, E. (2014). Making the case for space: The effect of learning spaces on teaching and learning. Curriculum and Teaching, 29(1), 5-19.
19. Yang, Q. (2019). Research on Active Service Model with Personalized Education Resources. Asian Journal of Contemporary Education, 3(1), 95-104.
20. Corbin, H. J., Chandran, D., Van Wingerden, C., & Baker-Sennett, J. (2019). The learning camera: A personalized learning model for online pedagogy in human services education. Journal of Technology in Human Services, 37(4), 334-346.
21. Bendahmane, M., El Falaki, B., & Benattou, M. (2019). Toward a personalized learning path through a services-oriented approach. International Journal of Emerging Technologies in Learning (Online), 14(15), 52.
22. Su, Y., Dong, Z., & Li, S. (2025). Distributed Sharing and Personalized Recommendation System of College Preschool Education Resources Under the Intelligent Education Cloud Platform Environment. International Journal of High Speed Electronics and Systems, 2540430.
23. Hu, J. (2023). Analysis of the feasibility and advantages of using big data technology for English translation. Soft Computing, 27(16), 11755-11766.
24. Wang, X. Y. (2022). Intelligent English translation and optimization based on big data model. Journal of Sensors, 2022(1), 8318921.
25. Lei, L., & Wang, H. (2022). Design and analysis of English intelligent translation system based on internet of things and big data model. Computational intelligence and neuroscience, 2022(1), 6788813.
26. Zhu, Q. (2023). Empowering language learning through IoT and big data: an innovative English translation approach. Soft Computing, 27(17), 12725-12740.
27. Liu, H. (2021, May). English translation flipped classroom teaching model based on big data. In 2021 2nd International Conference on Computers, Information Processing and Advanced Education (pp. 1215-1218).
28. Yan, S. (2024). Research on English Translation Teaching Model in Colleges and Universities Based on Big Data Platform. Frontiers in Educational Research, 7(8).

29. Wang, Y. (2022). English Translation Teaching Algorithm in Colleges and Universities Using Data-Driven Deep Learning. Mobile Information Systems, 2022(1), 2712541.
30. Yang, C. (2020, December). A study of influences of big data on machine translation and enlightenment for translation teaching in cross-cultural communication. In 2020 International Conference on Information Science and Education (ICISE-IE) (pp. 228-232). IEEE.
31. Cheng, W. (2017). Translation and Big Data Technology: Challenges and Implications. Lebende Sprachen, 62(2), 241-252.