# A Study of the Application of Computer Translation Technology in the Process of English Teaching

**Jianguo Liu, M.Sc**
*Jianguo Liu, Lecturer, Foreign Languages School, Henan University of Science and Technology, Luoyang, Henan, 471023, China*
*Correspondence LecLiu;hwlrlx@163.com*

**Objectives: With the development of Internet and computer technology, auxiliary teaching has been widely used in English teaching. Methods: Based on this, this paper studied English teaching based on the computer corpus. First, we designed and built the corpus based on the outline of the current English Teaching in China. Results:Secondly, the diagnosis algorithm was designed according to the needs of auxiliary teaching. And the implementation steps were expounded. Finally, we applied the algorithm designed in this paper to the evaluation of English composition.Conclusion: Through the test of the algorithm, the difference range between the test data and the actual data was acceptable. The algorithm designed in this paper was effective, and it could be further promoted in English teaching.**

**Keywords: computer; translation technology; diagnostic algorithm**
***Tob Regul Sci.™ 2021;7(5-2):4693-4701***
***DOI: doi.org/10.18001/TRS.7.5.2.34***

The applied talents are relative to the theoretical type (academic type and subject type), and there are differences in the two types. The former emphasizes applied knowledge, while the latter emphasizes theoretical knowledge. The former emphasizes the application of technology, while the latter emphasizes scientific research. The former emphasizes "special, fine, practical", and the latter emphasizes "wide bore, thick foundation" [1]. First of all, applied English teaching requires students to have a solid foundation of Chinese and English language, and have extensive knowledge of popular science culture [2]. English teaching requires not only the students to be familiar with the language and culture of the nation, but also to have a deeper understanding of the cultural background of the English language [3]. Secondly, applied English teaching requires students to have the ability to combine English knowledge with professional skills to practice. It is to require students not only to master technical knowledge, but also to combine the professional knowledge and the English language. It can convert the knowledge to the mother tongue and English, and translate its professional skills into English information, or combing the cost of national language in English. Apply what you learn to practice. Finally, the teaching of Applied English requires the students to have the quality of innovation. A solid language and culture is the basis, and professional knowledge is the necessary condition for the survival of the society. The quality of innovation is the necessary guarantee for success. In the teaching process of Applied English, the teachers should take the students as the main body and give full play to the students' subjective initiative. Teachers should mobilize students' interest and guide students to find problems, research problems, analyze problems and solve problems, and cultivate their curiosity and creativity [4].

At present, the teaching of translation is still in a relatively primary stage. Translation methods usually focus on the translation of words, phrases, sentences and the analysis of the grammatical structure. By doing a lot of exercises to improve the students' translation skills and abilities [5]. This translation teaching method ignores the

connotation and cultural characteristics of knowledge. It does not integrate language, culture and professional knowledge organically [6]. At the same time, the English environment is poor. There is not enough language environment and communication atmosphere. Students only carry out simple exercises and exchanges in class. After class, they only look at English movies and magazines. Students can hardly internalize their mechanically memorizing knowledge into their own knowledge. At present, English translation teaching does not combine language with actual communication environment and students' professional characteristics, but it pays little attention to students' translation psychology. There are two kinds of teaching modes of computer-aided translation technology: one is additional activity type, which are students learn computer assisted translation technology mainly through the network environment outside class. Teachers are regularly instructing. One is the combination of the traditional teaching method and the computer aided translation technology. It is taught by teachers to make computer aided translation technical learning programs to guide, supervise and evaluate students' learning results. Teachers use the network resources to set up the target course reasonably. When choosing network resources, it is good to consider whether the selected material is helpful to improve the students' translation skills and to what extent. The teacher should formulate a syllabus for computer aided translation techniques. Including theme selection, classroom organization, teaching methods, students' English level assessment, etc.[7].

## METHODS

### The Design and Construction of Computer Corpus

For the corpus, it is stored in the actual language materials that have appeared in the actual use of the language. Corpus is the basic resource that carries the language knowledge with the electronic computer as the carrier. The real corpus needs to be processed (analysis and processing) to become a useful resource. The assessment of the true level of the students' English is not only through the cold numbers of the scores, but also more humanized. This is the grammatical knowledge and vocabulary in the syllabus. So the basic words of the outline and their deformability are introduced first through a specific program, such as table 1.

Table1
Vocabulary of English And Its Modification

| Word | Part of speech | Suffix | Plural nouns | The superlative degree of comparison | Other forms |
|---|---|---|---|---|---|
| name | n. | | names | | |
| is | v. | | | | isn't |
| to | prep. | | | | |
| meet | v. | | | | |

According to the actual situation of the above picture, we can define the database of vocabulary, so that we can get all its deformable forms by any vocabulary. Then it is the knowledge point corpus, and the English knowledge point system divides the English knowledge points into five big plates. They are phonetic knowledge, vocabulary knowledge, grammatical knowledge, communicative competence and cognitive ability. Each plate is subdivided into one or three levels to form a hierarchical structure of four level knowledge points. In addition, each language section of English learning and testing is also used in a way to prompt the topics involved in their

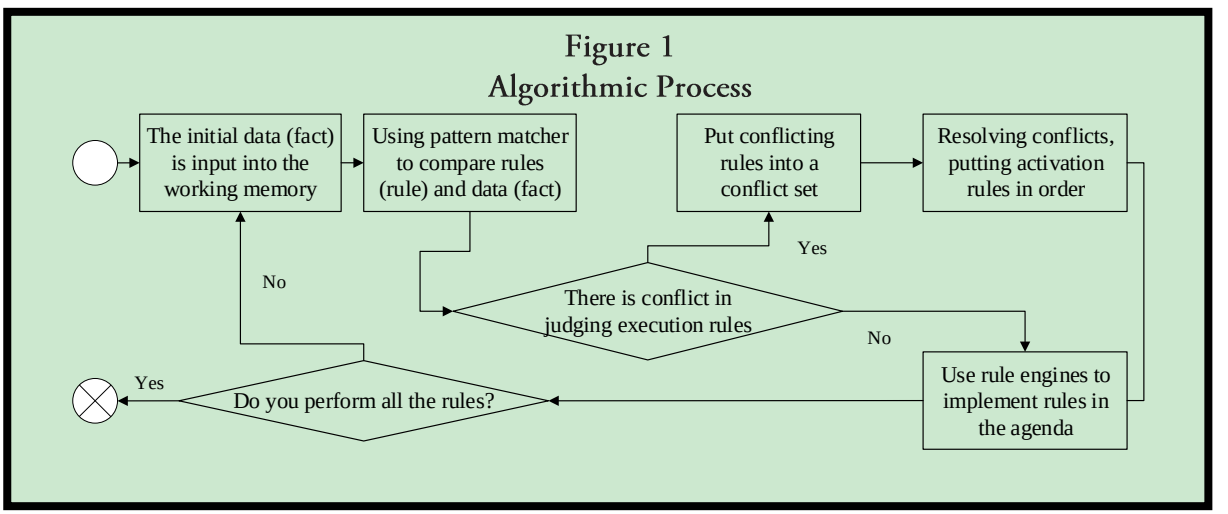
content. And call it the theme. It is also a part of the knowledge system. For every level of knowledge points, the basic rules are five parts: letter knowledge, L--- language knowledge, V--- vocabulary knowledge, G--- grammar knowledge, C--- communication ability, U--- cognition ability and T--- theme. The next level is expressed in two digits at each level. At most three levels, a total of 6 digits. For example, G01 represents the next level of nouns in grammatical knowledge. G0101 represents the function of nouns in the lower level of the nouns of grammatical knowledge. G010101 represents the function of nouns in the noun of grammatical knowledge as the subject of the next class of nouns. The database can be designed according to the above requirements.

If the existing knowledge point is not the parent - level knowledge point, then the corresponding p_ken_id is null value. If this is not the case, then the ID that corresponds to the parent level knowledge point is a one-dimensional data model, although it is reflected in the database. However, after the association of the father's knowledge points, a multi-dimensional tree structure can be embodied.

## Implementation of Computer Corpus Diagnosis Algorithm

We design the diagnostic algorithm. The first step is to initialize the rule engine. The rule engine has two ways of reasoning: the forward chain method and the reverse chain method. It corresponds to the two ways of thinking of human deduction and induction. The core algorithm is the Rete algorithm. The Rete algorithm is one of the most efficient algorithms used in the production system at present. At the same time, it is also the only decision support algorithm that has nothing to do with the number of execution rules. Therefore, it is more suitable for the corpus constructed in this article. Drools is an open source project based on this principle and implemented in the Java language. The reasoning step process of its rule engine is like the following figure 1:

Figure 1
Algorithmic Process




From the graph, we can learn that Drools encapsulates the Rete algorithm and adds the corresponding peripheral module extension function, which can provide support for many rules processing. It is also possible to achieve complex logic in the process of development. 7 conflict processing strategies are selected to deal with the conflict. One is the priority strategy. Any rule defines an integer priority. The rules with relatively high priority are precedence. If the priority is consistent, it is treated as a subset of the conflict rules to other conflict strategies. The second is the complexity priority strategy. In the case of the rule complexity, the more complex the rules are, the higher the complexity is, and the higher the priority is. The third is the simple

priority strategy. Contrary to the complexity priority strategy, this strategy takes the rules with the least condition and the lowest complexity to execute first. The fourth is the breadth strategy, which is added to the order of the agenda to deal with conflicts, and the rules that are added to the agenda can be preceded by the rules. The six is the loading number strategy. When each rule is added to the rule set, there is a specific loading serial number. The remaining unresolved rule subset is processed according to the loading sequence number. Seven random strategy belongs to the most simple strategy. The rules are randomly added to the agenda and are executed sequentially. The rule that is not sensitive to order is higher than that of Samsung. After a simple understanding of the Drools principle, it can be used to initialize the rules by virtue of it.

Then the key part is the calculation of sentence similarity. Text similarity is widely used. The rule is filtered by the threshold setting, and the regular matching is implemented by the composition rules in the association corpus. It can be oriented collocation, lexical structure, diction and related language ability to implement accurate judgment. Therefore, in the process of diagnosis of English composition, it is particularly critical to calculate the relevance of the sentence. In the previous studies, sentence correlation is also called sentence similarity. And it is defined as the number of [0 and 1] to represent the matching degree of matching on the semantic level of the two sentences. The larger the value is, the greater the similarity is between the two sentences. When the value is 1, the meaning of the two sentences is consistent. The smaller the value is, the lower the similarity is between the two sentences. When the value is 0, the semantics of the two sentences are completely inconsistent. This definition is not accurate enough, but it is easy to understand. In this paper, the text correlation is considered to be consistent with the meaning of similarity. In the process of Natural Language Processing, we use the TF-IDF method to judge two texts or sentence correlations. The relative entropy is a key concept in information theory. It is also called "cross entropy". It is specifically used to define whether two positive functions are similar or not. As far as two fully consistent functions are concerned, the relative entropy is zero. In the actual Natural Language Processing process, the relative entropy is usually used to define the two common words synonymous in grammar or semantics, or whether the contents of the two articles are consistent or not. Relative entropy can be used to obtain a particularly critical concept in information retrieval: the TF/IDF value.

TF is the number and normalization of the number of words in the sentence. It is with the total number of words divided by the number of times of keywords. It is also called "single text word frequency". For example, in 1000 of the total number of words, "look after", "my" and "Uncle" appeared 2, 35, and 5 times respectively. Then their word frequency is 0.002, 0.035, and 0.005, respectively. We add these three numbers, and the 0.042 is a simple measure of the correlation of the corresponding text and the sentence "look after my uncle". The abstract is expressed as if a query contains a keyword W1, W2,... Wn. Their frequency of word in a particular text is: TF1, TF2,... TFn. The correlation between a query sentence and the text is:

$$\sum_{i=1}^{n} TF_i \quad (1)$$

IDF: assuming that a keyword w appears in a Dw sentence, the larger the Dw, the smaller the weight of the w. Vice versa. Let D be the total sentence. The most used weight is the "inverse text frequency index", and its formula is:

$$\log \frac{D}{D_w} \quad (2)$$

In the above example, the word "my" accounts for more than 80% of the total word frequency, and it makes little sense for the purpose of determining the text. It is called "Stop words". It is not to take into account their frequency when measuring correlation. The words should be deleted as well as "the", "a", "an", "en" and so on. It can be analogous to the Chinese words, "the, ah, what," and so on. After ignoring the words that should be deleted, the similarity of the above text becomes 0.007. "Look after" contributed 0.002, and "Uncle" contributed 0.005. For a

marking, "the two word" look after "and" Uncle ", the importance is not the same. So we need to give each word a weight. The setting of this weight must meet the following two conditions. The greater the ability of a word to embody the theme is, the greater the weight is. On the other hand, the weight should be smaller, and the weight of the deleted words should be zero. Let's set up IDF1 , IDF2 , …… , IDFn is the weight of each keyword. Then the correlation calculation formula becomes the weighted sum [36] , that is:

$$\sum_{i=1}^{n} TF_i \times IDF_i \quad (3)$$

## RESULTS

In accordance with the needs of the research project, the paper is based on the three layered framework of J2EE technology. In order to facilitate the stability of the system and the effective use of hardware resources the entire environment is set up on a Linux service to deploy the database. Apply it to the server and on the WEB server. A class of 50 people was selected to carry out the overall test in the machine room. The fifth unit exam next semester junior middle school Wuluo road was selected as sample data. As far as standardized tests are concerned, in view of the fact that the actual diagnosis is relatively simple, it is not considered in the test. It is specific to consider the performance of the composition and the rate of error. A simple test is made for the following cases. The test steps include: students enter the established web site with the Internet and enter into a pre set examination room for a time limit answer. Any student needs to write a composition. After the arrival of the system, the system automatically submits the answer to the students. The diagnosis is operated with the background management interface, and the report is generated. The composition of the test paper is as specific as Figure 2 below:

Figure 2
Test Paper Topic

Soction
Please write a composition of about 150 words according to the following scenarios:
It's 8 o'clock in the evening, and your family is at home. According to the tips below, describe what everyone in your family is doing.
1) mother is washing clothes
2) father and brother play chess in the living room
3) grandpa and grandma watch TV together
4) I do my homework in my room

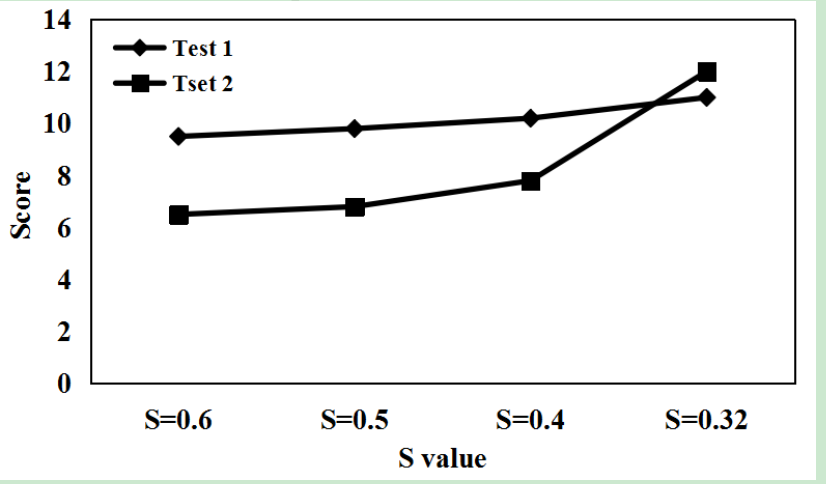
A rule can be generated by a tagging tool, such as a graph. In order to ensure the accuracy of the diagnosis as far as possible to improve the rule base, so a complete composition mark may be a lot. Then we have a personal diagnosis test. A single test was carried out from the test class to get the highest score and the lowest score. Student A's composition, the standard score is 13.5 points. Student B's composition, the standard score is 11.25 points. Student C's composition, the standard score is 12.5 points. Student D's composition, the standard score is 9 points. Student E's composition, the standard score is 5 points. The diagnostic results of the lower table were obtained by the diagnosis. S refers to the threshold of sentence similarity, and the corresponding results are scores and time (unit seconds) point only standard score, as shown in Table 2.

Table2
Diagnostic Results

| Student | point | S=0.6 | | S=0.5 | | S=0.3 | |
|---|---|---|---|---|---|---|---|
| A | 13.5 | 6.6 | 0.126 | 13.6 | 0.267 | 13.6 | 0.467 |
| B | 11.25 | 2.7 | 0.171 | 10.3 | 0.475 | 12.3 | 0.563 |
| C | 12.5 | 6.2 | 0.113 | 11.2 | 0.241 | 11.2 | 0.272 |
| D | 9 | 10.0 | 0.066 | 10.0 | 0.171 | 10.0 | 0.204 |
| E | 4.8 | 3.7 | 0.071 | 3.7 | 0.241 | 3.7 | 0.360 |

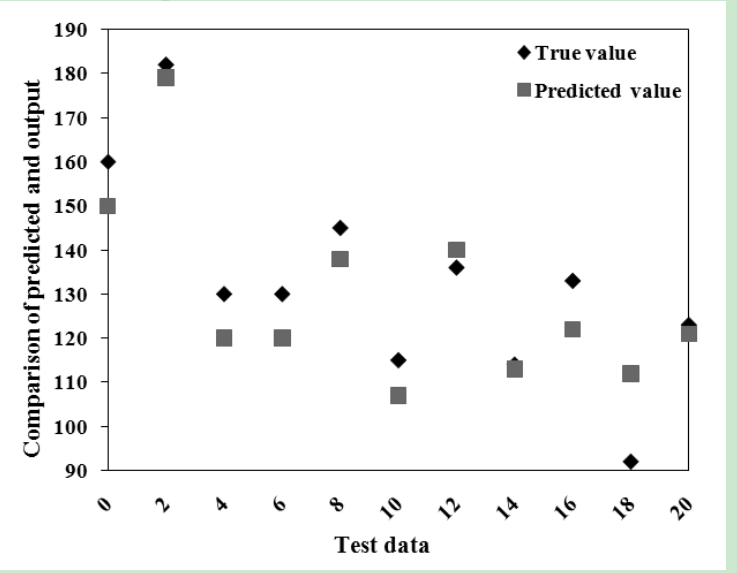
Based on the repeated tests of the upper table test data, we find the following results: when the similarity threshold is reduced, it will make the rules matching with students continue to increase. The number of calculations that correspond to the number of positive matches and the minimum edit distance will continue to increase. In this case, the calculation time will increase significantly. Only in terms of large-scale applications, performance is bound to be significantly reduced. It must define the accuracy and timeliness of the composition score. Only in the Internet application, the user experience is particularly critical. So it is crucial to select the appropriate threshold. For a composition with low score, the similarity of the sentence is less than the threshold in view of the less rules of the rule base. So the adjustment of the threshold does not have a great impact on it. The selection of threshold and composition score are basically a few. The threshold selection is within the 0.2-0.3 interval. The error rates of students A, B, C, D and E were 0.67%, 7%, 8.67%, 6.67% and 6.67% respectively. When it is known that the error rate is relatively large, it is necessary to repeat the new rules of the composition reference in order to prevent the problem of large error in diagnosis due to incomplete corpus.

Figure 3
Comparison of Class Scores




Then we combine the above steps to test the whole class and collect all the English compositions of all the people. It is tested by five front-line teachers. After obtaining five grade teachers' grading scale, the average score of 52 people in the class was 12.43, which was regarded as the marking value of the average score of class composition. The actual measured data are shown in Figure 3. In the case of the similarity threshold S=0.32, the overall error rate of the class composition is 9%. After testing the individual and class, the error rate is within 10% and the difference is within 2 points, which can meet the established design goals. It can be applied to practice. Figure 4 below is a comparison of the model test data built above and the actual data. It is known that the difference between the test data and the actual data can be controlled within two standard deviations. The algorithm designed in this paper is effective.

Figure 4
Comparison of Data Calculation Results




## DISCUSSION

Nowadays, the rapid development of computer technology and Internet technology made it gradually started to be applied in the teaching for English teaching, in view of its relative complexity. So computer aided was the key to the teaching of English. Based on this background, this paper focused on the implementation of ICAI English teaching. According to the selected syllabus, three levels of English anticipation were set up in this paper. It covered lexical knowledge, grammatical knowledge, communicative competence and cognitive competence modules. In the end, a class of auxiliary algorithms was designed in accordance with the needs of auxiliary teaching. On the basis of the TF-IDF algorithm, the evaluation of English sentences was implemented. Then we discussed the implementation of the algorithm. Finally, we applied the designed algorithm to English composition assessment, evaluated the composition of students in the case class, and compared it with the English teacher's manual test. It was concluded that the difference between the test data and the actual data was in the acceptable range by the test of the individual and the class. The validity of the algorithm was verified, and it could be popularized in English teaching.

## Human Subjects Approval Statement

This paper did not include human subjects.

## Conflict of Interest Disclosure Statement

None declared.

## References

1. James L M, Belitskaya-Lévy I, Lu Y, et al. Development and application of a diagnostic algorithm for posttraumatic stress disorder. *Psychiatry Research,* 2015, 231(1):1-7.
2. Lee H M, Kang M K. The effective teaching methods of modal auxiliary verbs through the analysis of middle school English textbooks. *Information Japan,* 2016, 19(12):5691-5696.
3. Neels J. The history of the quasi-auxiliary use(d) to: A usage-based account. *Journal of Historical Linguistics,* 2015, 5(2):177-234.
4. Perek F, Hilpert M. A distributional semantic approach to the periodization of change in the productivity of constructions. *International Journal of Corpus Linguistics,* 2017, 22(4):490-520.

5. Sharath B N, Shastri S. India's new TB diagnostic algorithm – far from reality?. *Public Health Action*, 2016, 6(3):206.
6. Tholpadi G, Bhattacharyya C, Shevade S. Corpus-Based Translation Induction in Indian Languages Using Auxiliary Language Corpora from Wikipedia. 2017, 16(3):1-25.
7. Wei X, Lin H, Yu Y, et al. Low-Resource Cross-Domain Product Review Sentiment Classification Based on a CNN with an Auxiliary Large-Scale Corpus. *Algorithms,* 2017, 10(3):81.