Deep Belief Networks Aided Real-time Business English Translation System

Wenli Xu[1,*] and Yanli Liu[2]

[1]International Education College, Zhengzhou University of Light Industry, Zhengzhou, Henan 450002, China, xuwenli@zzuli.edu.cn
[2]International Education College, Zhengzhou University of Light Industry, Zhengzhou, Henan 450002, China, liuyanli202210@163.com

Corresponding author: Wenli Xu, xuwenli@zzuli.edu.cn

Abstract. Business involves the rights and obligations of both parties in business activities, and requires that the style of words, sentences and chapters be formal and free of personal feelings. Therefore, we should try our best to translate the words with the same style in Chinese to reproduce the style of the original text. In order to avoid cultural differences, translators should try their best to match the meaning and connotation of the original text when translating with the help of functional equivalence theory. This paper studies the application of DL (Deep learning) in computer-aided real-time business English translation system. A computer-aided real-time business English translation system is established, and a temporal annotation algorithm for business English based on DBN (Deep Belief Network) is proposed. The temporal information of Chinese phrases is extracted by DBN. Then, according to the temporal information obtained by the temporal labeling algorithm, the verbs in the sentence are modified into the corresponding tense verbs. The research results show that on the BLEU (Bilingual Evaluation Underground) evaluation index, the integrated model in this paper can reach 88.186 when the fuzzy interval is set to the maximum value. Compared with 1418.054 characters/s, this system is 2945.105 characters/s, which is 1527.051 characters/s higher than that of another system. It can be seen that the translation speed of this system is faster and the system performance is superior.

Keywords: Deep Learning; Computer-Aided; Business English; Translation System.
DOI: https://doi.org/10.14733/cadaps.2023.S7.131-141

1 INTRODUCTION

More business information is presented, which gives more requirements to business English translation. English translation needs to achieve the same stimulating effect from the aspects of semantics, style and style. There are a lot of technical terms in business, which are professional and accurate [1]. There are many fixed sentence patterns in business English, especially in business letters, and most of them have fixed translation patterns and word order. Business English translation, as a practical course, urgently needs a mode that adapts to the development of the post and reflects the professional characteristics [2]. At present, there are still many problems portable instant translation system. For example, the existing echo cancellation algorithm often diverges in the case of two-terminal intercom, the convergence speed of the algorithm cannot meet the real-time requirements, and the interference of model noise is not considered, which leads to poor user experience [3].

The study of machine translation has been put forward since the birth of computer. It has experienced twists and turns, with success and excitement, but more frustration and confusion. The transformational generative grammar proposed by Zhu et al. has made great progress in linguistic theory, and machine translation began to focus on transformational grammar rules [4]. The traditional way of translation memory retrieval is string-based retrieval, and most translation memory systems are string-based retrieval. Zappone et al. put forward a method of integrating translation memory with statistical machine translation based on word alignment information, which aligns words in translation memory, extracts useful phrase rules from it, gives high scores and adds them to the phrase list [5]. With the wide use of computer - assisted English teaching, its convenience is gradually revealed. It is not only helpful in data processing. Wang et al. [6] believe teachers can also use the network function to randomly check students' learning achievements and obtain feedback information in a timely manner. The answers can also be sent to all students' computers for comparison. It is convenient for English teaching. Yao [7] The development of information technology and the application of computers have had a great impact on foreign language teaching. The teaching process of teachers and the learning process of students are closely integrated to promote teachers to create new ideas for teaching ideas and promote fundamental changes in the teaching process. This makes students change the traditional passive learning multimedia technology can be used to design a new overall teaching process and interactive, personalized training methods. Form a new combination of teachers, students, textbooks and teaching methods. Zhang [8] taking the English forum module as an example, computer assistance has played a great role in stimulating students to communicate in English. In the English forum discussion, students will forget their fears and try to communicate in English. Along with the changes, there are also classroom performances, and more and more students participate in classroom activities. Teachers should give appropriate encouragement and guide students to speak freely on topics they want to discuss. At the same time, teachers do not need to participate in the discussion, but supervise students' participation as bystanders. Students can not complete communication because of language anxiety and different levels of language ability. In the forum module, students can choose different topics for communication and discussion in English according to their own interests and levels.

Through a brief overview of the definition and core technology of CAT, Zhou and Gao [9] Computer aided systems do not encourage single handed learning or teaching. It shares common interests and needs among a group of like-minded people by creating a positive learning and teaching environment. Its effectiveness lies in the participation and interaction of learners. Through this new model of English teaching and online activities, students show a positive learning attitude. Even in offline activities, the communication between students and teachers and classmates is more successful. In completing oral reports in online tasks, students tend to achieve better results, especially in building reports and using new vocabulary. Zhou and Ma [10] think this problem can be solved by using computer - assisted multimedia to carry out individualized teaching. For example, in the process of making courseware, the customized animation in each step can be used to encourage students and improve their self-confidence with simulated applause. They can also compile their own conversation content to give them the opportunity to fully display their talents and further improve their communicative competence. The students with poor learning level are only required to complete any group of conversations according to the content of the text. For example, in the previous conversation, I asked the better students to complete several groups of dialogue exercises. Business involves the rights and obligations of both parties in business activities, and requires that the style of words, sentences and chapters be formal and free of personal feelings. Therefore, we should try our best to translate the words with the same style in Chinese to reproduce the style of the original text. In order to avoid cultural differences, translators should try their best to match the meaning and connotation of the original text when translating with the help of functional equivalence theory. Therefore, if we can't consider the semantics and connotation at the same time, we need to use the form change to present the corresponding semantics and cultural characteristics of the original text. Through the analysis, it is found that the oral interaction problem can be regarded as a transient interactive learning problem, so it can be solved by adopting DL (Deep learning) algorithm. No matter the size of the data, there will always be popular words or new words in the actual translation, so to solve this problem, we need to train the corpus at character level before the model training, and deal with vocabulary translation problems outside the vocabulary list.

2 RESEARCH METHOD

2.1 System Requirement Analysis

How to provide software, whether the market needs the system, where the funds to use and develop the system come from, and whether the hardware equipment is needed to use the system, etc., before we can find out whether the system is worth developing. A series of new development technologies have been catalyzed, such as java, C #, Python and a series of formed development technologies. Today, the technology of "speech recognition" is gradually developing, and the accuracy of speech recognition for people with large vocabulary can reach 98%. The accuracy of speech recognition for a specific person can be higher than that for a non-specific person. The database of this system is based on PostgreSQL database. The administrator can directly add, delete, and check the database of recorded information on the page. The operation is simple and clear, so the system meets the operational feasibility. As shown in Figure 1.

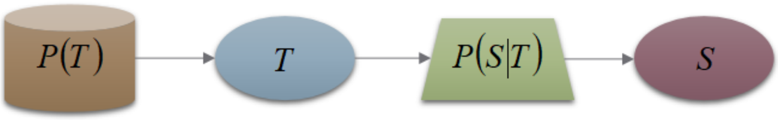

Figure 1: Noise channel model.

The probability value can be obtained according to Bayes formula:

$$P(T|S) = \frac{P(T)P(S|T)}{P(S)} \tag{1}$$

Among them, $P(T)$ is the probability that the target language text T appears, which is called language model, and $P(S|T)$is the probability that the target language text T is translated into the source language text S .

Language model is an important model in machine translation, which is used to evaluate the fluency of the translation.

$$p(w_i|history) = p(w_i|w_{i-1}, \cdots, w_{i-n+1}) \tag{2}$$

To evaluate the quality of machine translation, we can turn to the annotators to judge the quality of the translation, The most widely used automatic evaluation index of machine translation effect is BLEU (Bilingual Evaluation Under Study), which can take into account the problem of order when evaluating machine translation effect.

In n-gram, BLEU index is defined as:

$$BLEU_n = brevity\_penalty \exp \sum_{i=1}^{n} \lambda_i precision_i \tag{3}$$

$$brevity\_penalty = \min\left(1, \frac{output\_length}{reference\_length}\right) \tag{4}$$

Usually, the maximum order n of n-gram is set to 4, so this index is also called BLEU-4. In addition, the weight λ_i of different accuracy is set to 1, so the calculation formula of $BLEU_4$ is simplified as:

$$BLEU_4 = \min\left(1, \frac{output\_length}{reference\_length}\right) \prod_{i=1}^{4} precision_i \tag{5}$$

2.2 Overall System Design

There are different types of business texts in English, such as business letters, business contracts, insurance policies, letters of credit, etc., and each type has a fixed format, with each part explaining a certain key issue. The corresponding format should also be used when translating, and adjustments can be made according to the different expression habits of the two languages when necessary. Business English translation pays attention to the transmission of information. The two principles that business English translation should follow are accuracy and professionalism.

Artificial feature transformation or the design of nonlinear kernel function requires a lot of domain knowledge, and even it is difficult to find effective methods on some tasks. The most commonly used deep learning structure is deep neural network. It can be said that the current upsurge of research on neural network is mainly caused by its good experimental results in various artificial intelligence tasks; Most researches focus on designing a suitable deep neural network structure for a specific task.

One of which is the explicit layer h and the other is the hidden layer v. For RBM(Ristricted Boltzmann Machine), the following special parametric representations are most commonly used:

$$E(v,h,W,a,b) = -\sum_i \sum_j w_{ij} v_i h_j - \frac{1}{2} \sum_i (v_i - b_i)^2 - \sum_j a_i h_j \tag{6}$$

The characteristic of RBM is that the conditional probability of another variable is easy to calculate when one variable is given.

In Transformer, both position coding and word embedding are dmodel dimensions, so they can be added directly. The choice of location coding is not fixed, and different researchers put forward different location coding methods. There are both learned and fixed ones. Trigonometric function is used in Transformer for position coding:

$$PE_{(pos,2i)} = \sin\left(\frac{10000^{\frac{2i}{d_{model}}}}{pos}\right) \tag{7}$$

$$PE_{(pos,2i+1)} = \cos\left(\frac{10000^{\frac{2i}{d_{model}}}}{pos}\right) \tag{8}$$

pos here is the position, and i is the dimension.

In this system, the client requires multiple terminals, and the mobile terminal and PC terminal should be supported. The mobile terminal is developed under the Android environment, and the Android operating system can run not only on mobile phones, but also on tablets or other smart devices. The PC is developed based on B/S architecture, the program is easy to maintain and update, and the compatibility of the system does not need to be considered. The system application architecture is shown in Figure 2:

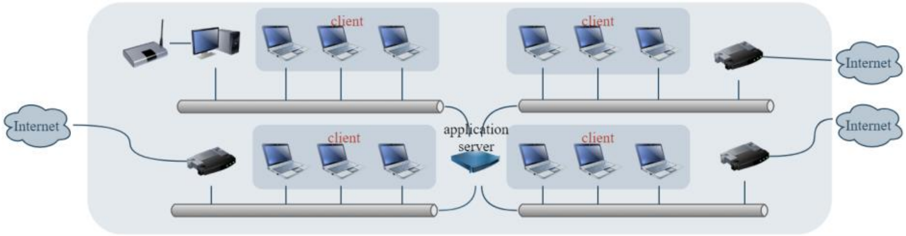

Figure 2: System application architecture diagram.

The realization of the translation function is to use the speech recognition function to first convert the speech into the text of the same language, and then translate the text into the text of the specified language through the machine translation interface provided by the software.

2.3 Key Technology Realization of the System

Business English is the language used in international trade, and it is the basis for both sides to reach economic cooperation and business trade. Translators should mark this, so that the information receiver can easily receive the information sender's signal or the conference topic, and the information receiver can judge the signal released by the other party to avoid unnecessary trouble. In business English translation, translators should pay attention to the overall structure of sentences, and combine the context before and after to translate information, so as to avoid the phenomenon of taking it out of context. Among them, the professional content or vocabulary should be expressed and paraphrased in professional terms, and the information you are not sure about should be prepared in advance. Translators translate and paraphrase by doing their homework in advance or combining with the background of the times and society, so as to avoid the wrong transmission of the information received by the information receiver.

DBN is used in business temporal tagging algorithm. Different from the traditional discriminant model. At present, there are many existing machine translation models, such as OpenNMT and Transformer. This paper adopts the Transformer translation model based on attention mechanism, which has the advantages of high translation performance and fast training speed. Figure 3 is the overall process of computer-aided real-time business English translation system.

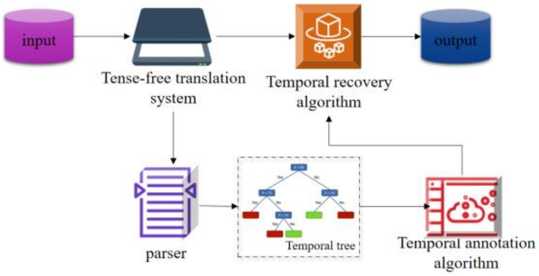

Figure 3: The overall process of computer-aided real-time business English translation system.

Firstly, this paper analyzes the input Chinese sentences and obtains the Chinese syntactic analysis tree. Secondly, the Chinese syntactic analysis tree is constructed into Chinese incomplete temporal tree, and the Chinese complete temporal tree is obtained by temporal annotation of Chinese incomplete temporal tree using Business English temporal annotation algorithm. The English verbs are transformed into verb forms with corresponding tenses. This has the advantage that the translation system doesn't need to focus on tense processing, but can focus more on the semantic translation of sentences.

Assuming that the tree to be labeled has N internal nodes, and the number of labeling sets of each internal node is M, there are $M^{'}$ labeling situations in the tree to be labeled. From all the annotation situations, one annotation with the highest generation probability is selected as the whole annotation of the tree to be annotated, and this whole structure is called tree annotation model in this paper. The input of the tree annotation model is an incomplete annotation tree, and the purpose is to get the best annotation of the tree. The best annotation can be expressed in mathematical form as follows:

$$\hat{x}_1, \hat{x}_2, \cdots, \hat{x}_N = \underset{x_1, \cdots, x_N \in H}{\arg\max} P(IT[X]) \tag{9}$$

Where N is the number of internal nodes, H is the label set of internal nodes, and x is the label of all internal nodes, that is, X .

Temporal recovery algorithm is an operation after obtaining the non-temporal English sentence, the alignment relationship between Chinese and English, and the marked Chinese temporal tree. In negative sentences, the negative word "not" needs to be taken into account in the morphological transformation of English verbs. For other cases, it is only necessary to use the rules to find the corresponding tense verbs from the verb morphology table and replace them.

The impulse response time of the echo path is very long, but the order of the common parameter adjustable transverse filter is shorter than that of the real echo path, so the model noise is ubiquitous in practical applications, and it is necessary to quantify and process the model noise. The finite impulse response of that echo path and the finite impulse response of the filt are defined as:

$$h = [h_0, h_1, \cdots, h_{N-1}, h_N, h_{N+1}, \cdots, h_{L-1}]^T \tag{10}$$

$$\hat{h}(t) = [\hat{h}_0(t), \hat{h}_1(t), \cdots, \hat{h}_{N-1}(t)]^T \tag{11}$$

Where L is the order of the echo path; N is the order of the transverse filter with adjustable parameters, and $N < L$.

Considering that the division of short sentences is usually based on commas, in the design training project, commas are still used as the identification marks of short sentence features of translation learning samples, and VDQ is defined as the full stop identification mark of short sentence training samples. VDV represents the comma identification string identification of the short sentence training sample; It can be obtained that the short sentence training feature function of the training sample is:

$$VDQ = VDV_1, VDV_2, \cdots, VDV_n \tag{12}$$

In the training process, if the logo string is complete, the position representing VDV is the comma cutting position of the short sentence. After the short sentence is cut, mark the cutting process with "+"; If the above conditions are not met, the training will automatically skip the VDV mark and start the segmentation and recognition training of the next short sentence.

In the process of coding-decoding machine translation, the first translated word will affect the translation of the latter word. The decoder model based on the attenuation weight loss function is to give higher weight to the first word and lower weight to the later word in the decoding process.

In the process of translation, the decoding process of a word in the target language by the decoder can be understood as a simple neural network. Assuming that the input information required for decoding in this step is x, the loss function in the translation of this word is calculated as follows:

$$C_i(\theta) = \frac{1}{2}\sum_{x_i}\left\| y(x_i) - a \right\|^2 \tag{13}$$

Where i is the index order of the words to be translated in the target language sentence, y is the decoding unit, and a is the real output.

Entries and weights can be used as coordinate axes and coordinate values of an n-dimensional vector space, and each point in the space corresponds to the mapping of documents. Obviously, the vector angle is inversely proportional to the similarity. The cosine of the angle between vectors is used to calculate the similarity, and the formula is as follows:

$$sim(Q, D) = \cos\theta = \frac{\sum_{i=1}^{m}(q_i \times d_i)}{\sqrt{\left(\sum_{i=1}^{m} q_i^2\right)\sum_{i=1}^{m} d_i^2}} \tag{14}$$

Therefore, sentence similarity based on vector space model can be obtained.

We think that the difference of hidden expression h^E, h^C should be as small as possible, so as to reduce the difference between Chinese and English vectors of the sample. Therefore, defining bilingual semantic loss L_{sem} reduces the gap between the two languages, as shown in formula (15).

$$L_{sem} = \left\| h^E - h^C \right\|_2^2 \tag{15}$$

Finally, this paper calculates the weighted sum of classification loss and bilingual semantic loss as the objective function of the whole model, and learns the model parameters based on Ada Delta algorithm.

$$L = \alpha \cdot L_{pre} + (1 - \alpha) \cdot L_{sem} \tag{16}$$

Where α is the weight of classification loss.

3 ANALYSIS AND DISCUSSION OF RESULTS

In the task of cross-language hedges recognition, about 2G Chinese corpus was downloaded, involving news, Weibo and other fields. Word2vec tool was used to train 25, 50, 100 and 150-dimensional word expressions.

Through the above experimental settings and related descriptions, and the results obtained are compared with the baseline translation system as shown in Table 1 and Table 2, and Figure 4 and Figure 5.

Model	*Development set*	*Test set*
Baseline	42.235	41.595
DeTense1	41.622	41.666
DeTense2	41.241	40.947
our	40.208	40.158

Table 1: Comparison of translation effects of the last model.

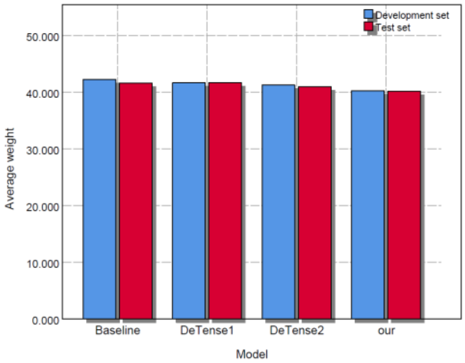

Figure 4: Comparison of translation effects of the last model.

Model	*Development set*	*Test set*
Baseline	42.74	41.399
DeTense1	42.505	43.926
DeTense2	43.525	41.776
our	41.975	41.806

Table 2: Comparison of translation effects of the last 10 models on average.

Baseline is a model trained by using raw data for Transformer model, and its super parameters are consistent with those of the non-temporal translation system.

In view of the first experiment, that is, the experiment of machine translation in Britain and Germany, this paper adopts the method of calculating bleu score to measure the translation effect. Through these iterations, it can be seen that the improved method can improve the task training rate. Figure 6 shows the experimental results of English named entity recognition.

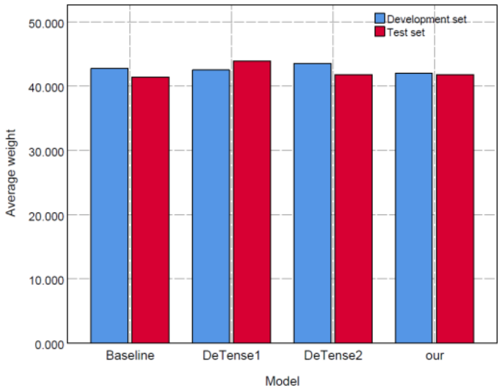

Figure 5: Comparison of translation effects of the last 10 models on average.

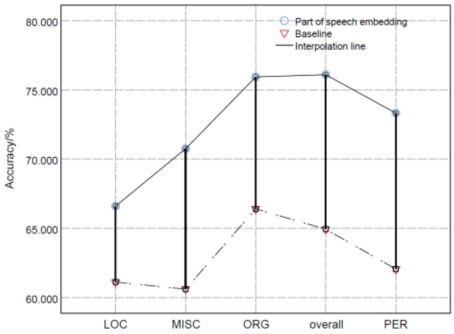

Figure 6: Experimental results of English named entity recognition.

It can greatly improve the convergence efficiency of translation training. Therefore, for other tasks such as machine translation, we can try to decide the division of groups according to specific usage scenarios and linguistic features, such as the aforementioned synonyms, which may be more useful for improving the effect.

Because the integration in this paper is not directly used as a translation result, but participates in the following statistical machine translation process, it cannot be guaranteed that the translation unit is the final translation result of the system. In order to observe the performance changes of each system more clearly, turn the results into graphs, as shown in Figure 7:

On the evaluation index of BLEU, the integrated model in this paper can reach 88.186 when the fuzzy interval is set to the maximum value, which fully reflects the influence of the retrieved translation memory fragments on the final translation results of the system. Under the condition of setting the similarity threshold reasonably, the performance of this system is still better than that of two benchmark systems.

And compare this system with the traditional machine translation system. The comparison results are shown in Table 3.

It can be seen that the translation speed of this system is 2,945.105 characters/second compared with 1,418.054 characters/second of the traditional machine translation system, which is 1,527.051 characters/second higher than that of another system. Therefore, the translation speed of this system is faster and the system performance is superior.

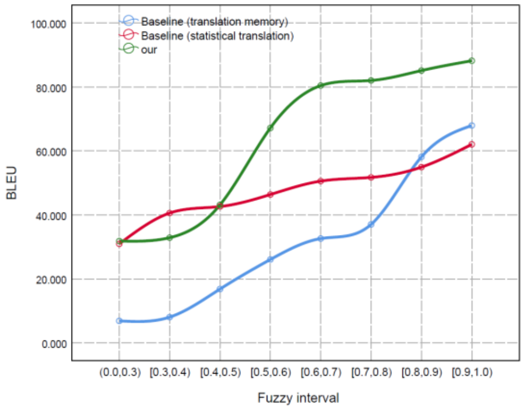

Figure 7: Comparison of different fuzzy interval systems.

System ☐	*Translation speed (characters/second)*
our	2945.105
Baseline (statistical translation)	2727.703
Baseline (translation memory)	1418.054

Table 3: Comparison of systematic translation speed.

Statistical machine translation system for phrases needs to integrate the translation model, language model, ordering model and other parts, and dynamically assemble the translation assumptions in the decoding part. Therefore, the translation speed is slower than that of translation memory system, and its decoding speed is also related to the length of the sentence to be translated. Therefore, it takes more time, which leads to the translation speed of the system being slightly worse than that of phrase statistical machine translation system.

4 CONCLUSION

There are many fixed sentence patterns in business English, especially in business letters, and most of them have fixed translation patterns and word order. Business English translation, as a practical course, urgently needs a mode that adapts to the development of the post and reflects the professional characteristics. Business English translation system is established, and a temporal annotation algorithm for business English based on DBN is proposed. Training 25, 50, 100 and 150-dimensional word representation with Word2vec tool. The integrated model in this paper can reach 88.186 when the fuzzy interval is set to the maximum value, which fully reflects the influence of the retrieved translation memory fragments on the final translation results of the system. Compared with the traditional machine translation system of 1418.054 characters/second, the translation speed of this system is 2945.105 characters/second, which is 1,527.051 characters/second higher than that of another system.

5 ACKNOWLEDGEMENTS

This work was supported in part by the Key Scientific and Technological Project of the Higher Education Institutions of Henan Province, China: Research on real-time early warning mechanism of intelligent video surveillance under big data (No. 22A520012).

Wenli Xu, https://orcid.org/0000-0002-9841-6420
Yanli Liu, https://orcid.org/0000-0003-3642-5409

REFERENCES

[1] Gao, Y.: Computer-aided instruction in college English teaching under the network environment, Computer-Aided Design and Applications, 18(S4), 2021, 141-151. https://doi.org/10.14733/cadaps.2021.S4.141-151
[2] Karpińska, P.: Computer Aided Translation–possibilities, limitations and changes in the field of professional translation, The Journal of Education, Culture, and Society, 8(2), 2017, 133-142. https://doi.org/10.15503/jecs20172.133.142
[3] Shu, Y.: Experimental data analysis of college English teaching based on computer multimedia technology, Computer-Aided Design and Applications, 17(S2), 2020, 46-56. https://doi.org/10.14733/cadaps.2020.S2.46-56
[4] Wang, B.; Ping, Y.: Perceptions of machine translation and computer-aided translation by professionals and the general public: A survey study based on articles in professional journals and in the media, International Journal of Translation, Interpretation, and Applied Linguistics (IJTIAL), 2(2), 2020, 1-14. https://doi.org/10.4018/IJTIAL.20200701.oa1
[5] Wang, B.: Empirical study on the computer-aided college English translation teaching, International Journal of Emerging Technologies in Learning (Online), 11(12), 2016, 68. https://doi.org/10.3991/ijet.v11i12.6012
[6] Wang, X.: Building a parallel corpus for english translation teaching based on computer-aided translation software, Computer-Aided Design and Applications, 18(S4), 2021, 175-185. https://doi.org/10.14733/cadaps.2021.S4.175-185
[7] Yao, S.: Application of computer-aided translation in English teaching, International Journal of Emerging Technologies in Learning (iJET), 12(8), 2017, 105-117. https://doi.org/10.3991/ijet.v12.108.7145
[8] Zhang, Q.: Computer Aided Translation Technology Application in ESP Teaching, Canadian Social Science, 12(8), 2016, 20-24. https://doi.org/10.3968/8739
[9] Zhou, W.; Gao, B.: Study on the application of computer-aided translation (CAT) in translation teaching, US-China Foreign Language, 14(12), 2016, 849-856. https://doi.org/10.17265/1539-8080/2016.12.004
[10] Zhou, X.; Ma, X.: Research on computer-aided translation, Scholars International Journal of Linguistics and Literature (SIJLL), 4(7), 2021, 213-215. https://doi.org/10.36348/sijll.2021.v04i07.003