Study on the Application of Fuzzy Algorithm in English-Chinese Translation System in Vocational College Education

Fen Han(✉)

Xi'an Vocational and Technical College, Xi'an 710077, Shannxi, China

Abstract. In E-C translation system, fuzzy algorithm can extract the best context and feature words and improve the accuracy of translation. This paper studies the English Chinese translation system based on fuzzy algorithm, and compares the translation effect of English Chinese translation system based on fuzzy algorithm. The simulation results show that, compared with the traditional E-C translation system, the E-C translation system using fuzzy algorithm has higher accuracy and recall rate, and higher translation accuracy.

Keywords: Fuzzy algorithm · English Chinese translation

1 Introduction

In recent years, with the deepening of globalization and the increasingly extensive international cooperation, English is becoming more and more important. As a tool of communication between English and Chinese, English Chinese translation system has become a research hotspot. There are two traditional English Chinese translation systems: rule-based and case-based. However, these systems can not extract the optimal context and feature semantics in the process of translation, resulting in the low accuracy of English Chinese translation. In addition, with more and more information at present, how to obtain the optimal translation from massive data becomes the key of English Chinese system [1]. Based on this, this paper proposes to introduce fuzzy algorithm into English Chinese translation system, and construct language mapping model by selecting optimal feature semantics, so as to achieve optimal translation of English Chinese translation system and improve translation accuracy. However, the commonly used fuzzy algorithm is easy to lead to the loss and lack of language, which affects the accuracy of translation. The improved fuzzy algorithm becomes the key to the research of English Chinese translation system.

2 Statistical Machine Translation

2.1 Introduction to Statistical Machine Translation

Machine translation is a very difficult job. According to the common understanding, to translate a sentence, we should first understand the source text, understand the meaning

M. A. Jan and F. Khan (Eds.): BigIoT-EDU 2021, LNICST 392, pp. 415–426, 2021.
https://doi.org/10.1007/978-3-030-87903-7_51

and sentence structure of each word in the source text, then select the translation of each word, and reorganize the sentence according to the grammatical structure of the translation. The traditional rule-based machine translation (RBMT) method is usually written by human language experts in various forms of rules, and then the computer completes the operation according to the rules written by human. This approach is effective to some extent, but when the number of rules reaches a large scale, the system becomes difficult to control. Moreover, in the face of all kinds of complex language phenomena, it is difficult to cover all the situations with manually written rules. Therefore, the rule-based machine translation system is often difficult to continue to improve after reaching a certain level [2–4]. Therefore, in early 1990, IBM company proposed a new machine translation method, called statistical machine translation method. The basic idea of this method is to collect a large number of bilingual sentences and the size of sentence pairs, and then use some algorithm to automatically learn translation knowledge from these corpora, and finally use these knowledge to translate, usually reaching 1 million.

2.2 Advantages of Statistical Machine Translation

2.2.1 Statistical Machine Translation Can Make Full Use of the Existing Translation

The basic working method of statistical machine translation is to establish statistical model, formulate algorithm rules, and use a large number of high-quality bilingual parallel idioms to train the system. In today's Internet era, there are a large number of bilingual texts accessed in home machine-readable format. After aligning and processing the bilingual texts, the statistical machine translation engine 18 can read and learn them, thus improving the construction efficiency and translation quality. For example, coge company imported 20 billion words of official documents from the United Nations into its translation engine for training, thus improving the quality of the translation engine in a relatively short period of time, making the translation engine of big Google able to catch up with others and ranking among the best in the evaluation RO of NST. have to Google's translation engine can be ranked in. The evaluation in NIST. They are at the top of the list.

2.2.2 Statistical Machine Translation has Higher Translation Quality

The traditional working mechanism of rule-based machine translation is realized through the in-depth study and continuous revision of language rules, and the machine transformation at the lexical, syntactic and other levels. The research of this transformation is complex and arduous, but the transformation result is stiff and unnatural, which can not be compared with manual translation. Statistical machine translation (SMT) is a new way to train students with high-quality manual translation, which makes the corpus in the translation system more authentic and more in line with the norms of manual translation. The earliest statistical translation system is based on the word translation model, which only considers the linear relationship between words, without considering the sentence structure and context information [5, 6]. When the word order of two languages is quite different, the effect is not good. With the continuous efforts of researchers, a phrase based translation model has emerged. The basic idea is to extract a large number of aligned

phrase fragments from large-scale corpus, and use these phrase fragments to match and combine the sentences to be translated. Due to the restriction of phrases, the choice of words is more accurate, and it is helpful to the translation of some common words and idioms, and the word order of translation is relatively determined, which makes the translation results more in line with the characteristics of the target language.

2.2.3 Higher Construction Efficiency and Translation Quality

The traditional rule-based machine translation (RBMT) is more complex in the study of language rules, which requires long-term work, huge development costs and repeated debugging. The representative of this aspect is the American SYSTRAN translation engine. Compared with RBMT, the construction and training of statistical machine translation system can be completed in weeks or even days, which greatly saves time and cost and improves development efficiency. Experts believe that if the same amount of money is invested in the machine translation system, the translation quality of the statistical machine translation system will be higher. (ER, 200518-21) Candide, a system developed by IBM in a short period of time by using statistical machine translation method, has a performance comparable to that developed by the famous machine translation company SYSTRAN after decades of development [7]. At present, the Institute of computing of Chinese Academy of Sciences is cooperating with Harbin Institute of technology and Xiamen University to develop SLK road statistical machine translation system. It is believed that it will make remarkable achievements in the near future.

2.3 Static Binary Compiler

FX! 32 contains the following interactive components: agent (transparent agent), loader (loader), runtime and emulator (runtime and emulator), binary translator (translator), server (server) and manager (Manager). Agent belongs to Windows NT process, which is used to detect whether there are x86 programs that need to be loaded and executed. If a program is detected to be executed, the loader is called to load the x86 code [8–10]. If the source program is executed for the first time, the emulator interprets the x86 code and gets its profiling information. Profiling contains the information of the static translator: the address of the call instruction to be interpreted, the source/destination address pair to jump indirectly, and the instruction address to perform the non aligned memory access. After the above program is terminated, FX! 32 server calls static binary translator and applies the collected profile information for static translation. At the same time, the generated profile information is stored in the database for subsequent execution of similar programs. Using this method, the source x86/WinNT program is translated to alpha machine incrementally [11]. The main drawback of FX! 32 system is that the first time execution is very slow, and it relies on pure static profiling, and can not dynamically optimize the code according to the change of program execution.

3 Fuzzy Algorithm

3.1 Fuzzy c-Means Clustering Algorithm

Fuzzy FCM algorithm is one of the most perfect algorithms in theory, and it is also very common in practical application. In essence, it belongs to the category based on partition algorithm, but it is a kind of flexible fuzzy partition, which introduces fuzzy theory on the basis of clustering, and then uses membership function to determine the classification of a data object.

An algorithm based on the set of FCM is a set of data $X = \{x_1, x_2, \ldots x_n\} \subset R^s$ is divided into C groups, and its cluster center is represented by C_i.

$$J(U, c_i, \ldots c_c) = \sum_{i=1}^{c} J_i = \sum_{j=1}^{n} u_{ij} d_{ij} \tag{1}$$

Where u_{ij} is between 0 and 1, C_i is the cluster center of the i class, $d_{ij} = \left\| c_i - x_j \right\|$ is the Euclidean distance between the ith cluster center C_i and the j-th data object x, and $m \in [1, +\infty)$ is a weighted index to control the fuzzy degree of clustering. The higher the m value, the greater the degree of ambiguity.

In order to improve the translation accuracy of English Chinese translation system, an English Chinese translation model system based on human interaction and feature extraction is proposed. Based on the overall qualitative understanding and logical reasoning ability of human-computer interaction, the translation similarity model is introduced into the English Chinese translation system model, and the translation results are obtained by calculating the translation similarity between two semantic loops in the same language space. Compared with SOA, SCA and SLA, the proposed algorithm has higher accuracy, accuracy, precision η and recall rate RecA, which provides a new method and approach for English translation.

3.2 The Concrete Steps of FCM Algorithm

The rule-based recognition method is mainly obtained from corpus automatically or compiled by experts, which has the advantages of easy to understand, but it is not universal, time-consuming and easy to produce ambiguity; the method based on statistics transforms the problem of noun phrase recognition into the problem of tagging similar words, which is simple and flexible, and does not rely on specific language model. It is a popular translation algorithm in the past; However, this method is based on a large number of sample data, which is prone to over fitting problems. With the rise of artificial intelligence and machine learning methods, neural network-based noun phrase recognition method has been applied to English noun phrase recognition. However, due to the complexity of English grammar, the accuracy of English noun phrase recognition needs to be improved. In order to improve the translation accuracy of English Chinese translation system, an English Chinese translation model system based on human interaction and feature extraction is proposed. Based on the qualitative understanding of human interaction and the ability of logical reasoning 3, the translation similarity

model is introduced into the English Chinese translation system model, and the translation results are obtained by calculating the translation similarity between two semantic vectors in the same semantic space. Compared with SOA, SCA and SLA, the English Chinese translation based on human-computer interaction and feature extraction has higher accuracy, accuracy, precision η and recall rate, which provides a new way for English translation. In semi supervised clustering, tag data acquisition needs a certain cost. Therefore, in the actual application of data sets, the amount of label data is relatively small, at this time SFCM algorithm can not reflect its advantages [12–14]. Therefore, this paper redefines the iterative formula of SFCM algorithm, and puts the parameters representing the weight of labeled data points in the iterative expression of cluster center, so as to adjust the influence of supervision information. Secondly, the iris data set in machine learning library is tested and verified by MATLAB programming. Finally, the accuracy of the improved ssfcm algorithm is analyzed and compared with the previous basic algorithm. The experimental results show that the improved algorithm has good performance.

4 Application of Computer Aided Translation Technology in Simultaneous Interpreting

This part introduces the application and main problems of computer assisted translation technology (hereinafter referred to as "technology") in the simultaneous interpreting of two stages, namely, pre preparation and on-site translation.

4.1 Technology Assisted Multimodal Pre Translation Preparation

Before the start of on-site translation activities, multi-modal pre translation with technical assistance can be carried out. The technologies used include online translation, machine translation, voice generation, etc. the technologies used include daofan translation, SDL studio17, Microsoft edge, etc. among the speakers of international conferences, some of them have strong accents, such as non-native speakers of English and native speakers of English speaking countries such as India and Singapore, which are often difficult to recognize. In the process of SDL studio17 automatic translation, you can choose to use Google translation, baidu translation, Sogou translation and other automatic translation engines for automatic translation, and choose online or local translation memory, Save translation results for future invocation. In addition, SDL Studio17, after the addition of multilingual control, the automatic generation of alphabetic terminology is easy to recite and find [15]. The simultaneous interpreting of medical conferences involving the author involves a lot of medical knowledge and terminology. It is difficult to learn and memorize in a short time. For this reason, the author consults relevant academic papers according to the speaker's information and the theme of his speech. Under the guidance of the concept of "multimodal interpretation preparation before translation", I watched the related topic speeches, classroom teaching videos and operation videos on domestic and foreign video websites, and analyzed the knowledge points, vocabulary, expression and sentence patterns. The Department also used the reading function of Microsoft edge browser and Baidu translation to input I English words into the browser, watch the words

and listen to the reading, In terms of the subject matter of the related topics of stimulating storage, the author found that most of the words, including terminology, are clearly pronounced in the Xiangya English Chinese Medical Dictionary and the Zhongshan medical dictionary, and most of the three pronunciations are very standardized.

4.2 Technology Assisted Live Simultaneous Interpreting Practice

Simultaneous interpreting is simultaneous interpreting. The translation is transmitted from the same device headset to the interpreter, and the translator is translators through the interpreter workbench microphone. Then, speech recognition technology, the translator's target language recognition into text, and projected on the front of the central screen. Speech recognition technology company arranges personnel on site to manually correct or delete Chinese words and sentences with serious recognition errors. All people at the meeting can choose to use headphones or watch subtitles to get the main idea of the speech. The translated words can be saved and used for meeting minutes, news reports, translation quality evaluation and control. It is worth noting that during the second session (2018) of the simultaneous interpreting, the speech recognition technology provided the company with the initiative to propose that the source language be transferred to the only display screen of the interpreter's workroom for reference. However, the author found that due to the excessive use of medical lingual and jargon, the effect of writing is not good [15]. Moreover, the simultaneous interpreting room and the conference hall are separated by walls, and can not directly observe the scene. If the only display screen is used for the display, it can not be used to broadcast the slide show on the spot. After weighing the pros and cons, the author and his partner agreed that the display screen should be used to play live slides. In the process of live translation, the author and his partner also used Google translation, Youdao translation and other online translation engines to translate unfamiliar words, phrases, sentences and paragraphs in the temporary speaker's Fantasy for reference. The author found that the translation and automatic pronunciation of most medical terms are very accurate. In the field translation, the author found that the intervention of technology has also brought some new problems, which can be summarized and analyzed from the perspective of speech recognition system and human factors.

5 Synonymous Translation

Synonymy is the most important relation in word. If the two expressions replace each other in the language text without changing their meaning, the two expressions are synonymous, and sy η set means synonymy [16–18]. Antonymy is not the basic organizational relationship of WordNet. It is a morphological relationship rather than a semantic relationship between concepts, such as the antonym pair heavy/light". As for the antonymy of adjectives, the semantic organization of descriptive adjectives is completely different from that of nouns. It adopts n-dimensional hyperspace structure instead of tree structure. Antonymy is the basic semantic relationship of descriptive adjectives, so synset of adjectives is connected through antonymy.

5.1 Technical Literacy and Psychological Quality of Interpreters in the Age of Technology Assisted Simultaneous Interpreting

In the context of technical assistance translation, the ability to use technology should be an important component of the quality of simultaneous interpreting. Technology, especially language technology, is a double-edged sword for translators. Various technologies help translators improve the quality of translation and the efficiency of pre translation preparation. For example, speech recognition can transcribe the target language produced by the interpreter, which is convenient for the interpreter to summarize, analyze and improve after translation. IFLYTEK was invited to attend the conference of simultaneous interpreting and interpreting at the end of 2018. At the meeting, iFLYTEK demonstrated its interpreter assistance software. This software can not only recognize the speech, but also highlight the numbers and proper terms in a special way, and provide the reference translation of terms. If this software constantly reduces the workload of simultaneous interpreting, it can even reduce the burden of memory and note taking. On the other hand, the mode of speech recognition and simultaneous interpreting has brought great psychological pressure to the interpreter. In the process of traditional simultaneous interpreting, the target language will only be kept in the brain in the form of voice [19, 20]. If the interpreter is careless, he will not be particularly careful. In the new mode of simultaneous interpreting and subtitle translation, the long time of omission in translation is displayed on the screen. This undoubtedly brings more psychological pressure to the interpreter. The interpreter needs better compression ability. In addition, after the introduction of speech recognition, the pronunciation is not standard enough and the effect of target language recognition is poor, which may be eliminated by the market. With the continuous progress of machine translation technology, a number of translators with poor translation quality have been pushed to the edge of unemployment.

5.2 How to Protect Intellectual Property Rights and "Innocence" of Interpreters in the Era of Technology Assisted Simultaneous Interpreting

Over the years, the protection of interpreters' knowledge has aroused great concern. Many kinds of intellectual property works of interpreters may be infringed and abused. For example, after the target language translated by the interpreter is recognized by the software, it will continue to be used by the organizer to add subtitles, write reports, and release technical materials. The excellent translation engine belongs to the secondary translation of the interpreter's knowledge, so that the vocabulary produced by the interpreter in the preparation of translation can be shared with the technology company and the host office for the high-quality and future translation quality. When I accept the task of translation, I agree with all parties about the intellectual property rights of the translation products, and the items and products that need to be paid for should be specified in the agreement. Except that the root cause of low translation is not always the poor level of the interpreter. Sometimes factors other than the interpreter, such as the on-site technology and physical environment of the speaker's speech, will affect the translation effect [21]. However, if the audience does not grasp the spoon (or pay for bilingual listening), they can not or have no time to consider and accurately analyze the reasons for the poor translation quality. Under the traditional simultaneous interpreting mode,

the translator will not only enhance the scapegoat transfer technology but also increase the omission of the interpreter. How to determine the responsibility of the interpreter becomes more complex. Translators should be strict with themselves and be good at using technical means to collect evidence and correct their names.

5.3 Technology Assisted Simultaneous Interpreting May Not Improve the Communication Effect of Speakers

Speech recognition and assisted simultaneous interpreting can help speakers generate speeches, so that they can be processed and edited in the later stage, and can be transmitted in text, or for the speakers or organizers to review and reflect after the meeting. However, it should be noted that some English speakers are not only very different from the typical standard native speakers in phonetic level, but also have great differences in the use of words, phrases, grammar, sentence patterns and text structures. The machine recognition of the corpus produced by them is difficult, and the effect of on-site translation and recognition is also poor. In addition, the use of simultaneous interpreting translation screen shows that this "information delivery" mode may lead speakers to be unable to attract audience attention. Voice transcribing technology providers_ A staff member described a typical scene to the author: a "heavyweight speaker is standing on the left end of the huge conference hall to make a report, and the live voice transcription screen is set on the right end of the conference hall. During the meeting, people pay attention to the right end of the screen, while the speaker is "in the cold", and the nonverbal communication with the audience is blocked.

6 An Analysis of English Chinese Translation System

6.1 System Architecture

The system mainly includes interface layer, implementation layer and data layer. The interface layer is mainly for the user to classify the translated text through the user interface; the implementation layer mainly completes the realization of each text classification process, designs the memory data storage, and provides the interface for calling the interface layer; the data layer is mainly used to complete the data exchange between the interface layer and the implementation layer. The details are shown in Fig. 1.

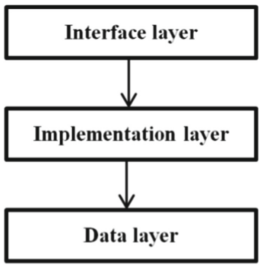

Fig. 1. System architecture of English Chinese translation system

In order to achieve user operation, the user interface is classified by MFC dialog box, which is divided into four modules: path selection, classification, preprocessing and classification change. TF-IDF and LDA are used for feature extraction. For the classifier design, according to the phenomenon that each category has unique feature words, the classification lexicon is designed. According to the frequency of each category of feature words in the document, the document category is first judged, and then other classification design is carried out by using Sm classification method [22]. According to formula (2), when Eq. (2) is satisfied, the document is directly set as the category, otherwise SVM classification is carried out.

$$\frac{Count(i)}{Total(N)} > 80\% \tag{2}$$

Where $Count(i)$ represents the number of feature words in the current Classified Thesaurus; $Total(N)$ represents the total number of feature words in the document set.

7 Analysis of Simulation Experiment

In order to ensure the accuracy and effectiveness of the analysis, the experimental parameters are set uniformly, including phrase translation amount of 500 characters, short text translation amount of 600 characters, translation rate of 15 kbps, semantic recognition rate of 25 kbps [23]. The only difference between simulation experiments of English Chinese translation system is that feature extraction algorithms are of different types. Tf-df feature extraction algorithm, LDA feature extraction algorithm and no feature extraction algorithm are used to process the preparation rate, recall rate and accuracy rate.

The system simulation results show that: compared with the English Chinese translation system without feature extraction algorithm, the τ f-df feature extraction algorithm has higher accuracy, recall and accuracy, and has higher translation accuracy; LDA feature extraction algorithm has higher accuracy, recall and accuracy, and better translation accuracy than τ f-df feature extraction algorithm. Theoretically speaking, the τ f-df feature extraction algorithm is realized by calculating document similarity when classifying documents. However, this method does not consider the semantics of feature words or sentences, resulting in two sentences with the same meaning or causal relationship being divided into different categories, and LDA feature extraction algorithm can identify [24]. Therefore, compared with the τ f-df feature extraction algorithm, LDA feature extraction algorithm has higher translation accuracy, which are shonw in Fig. 2 and 3.

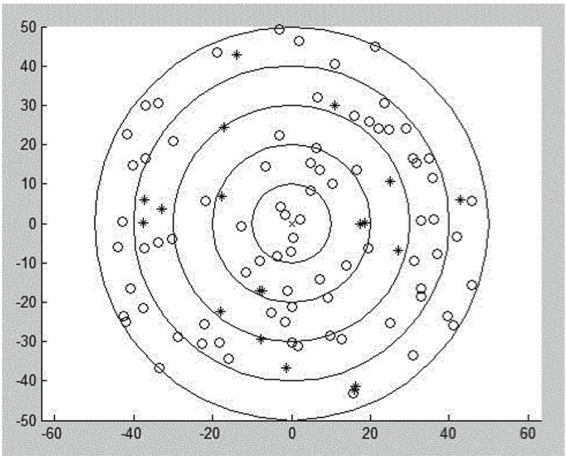

Fig. 2. Fuzzy algorithm with translation system

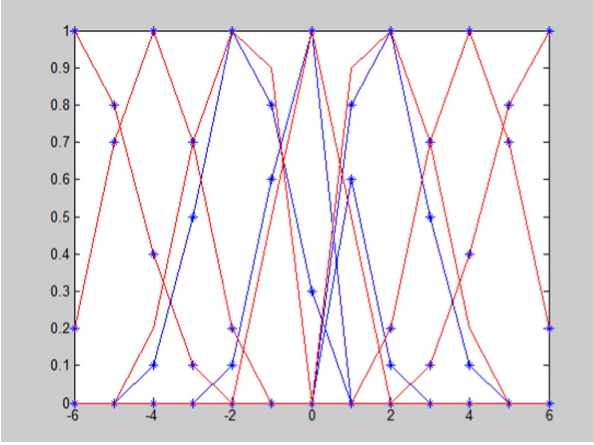

Fig. 3. Fuzzy evaluation simulation

8 Conclusion

This paper studies the English Chinese translation system based on feature extraction algorithm, and studies two types of feature extraction algorithms through simulation experiments. In the English Chinese translation system, feature extraction algorithm can be selected through feature semantic optimization, and construct semantic ontology mapping model to achieve the optimal solution and achieve the best English Chinese translation. Compared with τ f-idf feature extraction algorithm, LDA feature extraction algorithm has higher accuracy, recall and accuracy, and has higher translation accuracy.

Acknowledgements. 2020 Scientific and Research Program Funded by Shaanxi Provincial Education Department (20JK0366).

References

1. Ma, J., Huang, D., Liu, H., et al.: Research and application of English part of speech tagging for machine translation. China Commun. **3**, 66–75 (2012)
2. Wang, Y.: Part of speech tagging in English Chinese machine translation system. Comput. Eng. Appl. **46**(20), 99–102 (2010)
3. Zhang, Z., Yang, H., Zhao, Y., et al.: Topic text network construction method based on Gibbs sampling results. Comput. Eng. **43**(6), 150–157 (2017)
4. Guo, L.: Design of computer intelligent scoring system for English translation based on natural language processing. Mod. Electron. Technol. **42**(4), 158–160 (2019)
5. Fan, H.: Lack of subjective translation ability in translation teaching and countermeasures. Road Success **08**, 40–41 (2021)
6. Li, H.: On the importance of ethnic Chinese translation in cultural development. (05), 16–17 (2021). Huang He. Huang Huang Zi. Huang Zhong Ren
7. Song, D., Lin, Y.: Translation strategies of poetry and verse in the original version of a dream of Red Mansions. Chin. Cult. Stud. **01**, 144–157 (2021)
8. Li, X.: Legalization of local normative legal documents translation from the perspective of "rule of law". J. Nanyang Normal Univ. 1–6 (2021). http://kns.cnki.net/kcms/detail/41.1327.Z.20210311.1710.008.html
9. Li, Y.: The spread of Confucianism overseas under the influence of Confucian Classics Translation: a case study of Finglett's Confucianism research and its inheritance. Chin. Cult. Stud. **01**, 170–180 (2021)
10. Xi, W.: Translation strategies of conceptual metaphors in English and Chinese culture loaded words from a cross cultural perspective. Mod. Vocat. Educ. **11**, 98–99 (2021)
11. Li, Z., Wang, S.: Research on the cultivation of applied translation talents in nonferrous metallurgy – a review of metallurgical English. Nonferrous Met. (Smelting) **03**, 195 (2021)
12. Changshan: Translation strategies of modern agricultural science and technology English: a review of modern agricultural science and technology English listening and reading course. Plant Quar. **35**(02), 93 (2021)
13. Mao, H., Wu, L.: The red translator of Ningbo going out. Ningbo daily, March 2021. (009)
14. Wang, X., Zhang, J.: A study of "creative treason" in Chinese translation of Martin Fiyero. Sci. Educ. Wenhui (First Issue) **03**, 188–190 (2021)
15. Zhang, S.: Research on the ways to improve the core literacy of translators based on social needs. Foreign Lang. Teach. **42**(02), 55–59 (2021)
16. Geng, J.: Problems and countermeasures in English translation of imported and exported food packaging in China. Foreign Trade Pract. **03**, 59–62 (2021)
17. Liu, X., Wang, M., Zhang, Y.: The present situation and Prospect of translation studies on the biography of the arched hero – based on the analysis of CNKI literature resources. Cult. Ind. **07**, 60–61 (2021)
18. Zhang, Y.: The influence of cultural factors on subtitle translation from the perspective of Skopos theory: a case study of crazy animal City. Cult. Ind. **07**, 64–65 (2021)
19. Sang, D.: Research on garment English teaching and translation. Cotton Text. Technol. **49**(03), 94–95 (2021)
20. Palidan, A.: A study of conversion techniques and types in Chinese Uyghur translation. Media Forum **4**(05), 94–95 (2021)
21. Wen, Q., Guan, C.: A tentative exploration of intercultural communication in Translation Classroom Teaching: a review of pragmatics in Translation Classroom Teaching. Chin. J. Educ. **03**, 112 (2021)
22. Zhao, X.: A study of biblical translation by Chinese in the late Qing Dynasty and the Republic of China. J. Ningbo Univ. (Hum. Ed.) **34**(02), 8–16 (2021)

23. Sun, B.: Translation principles of Xixia officials and titles. J. Southwest Univ. Nat. (Hum. Soc. Sci.) **42**(03), 60–71 (2021)
24. Yan, D.: College English translation teaching based on intercultural thinking: a review of College English teaching and intercultural competence training. Chin. J. Educ. **03**, 131 (2021)