Discourse Analysis of Chinese Culture Translation Based on Crawler Algorithm

Rui Zhang(✉)

Xi'an Fanyi University, Xi'an 710105, Shaanxi, China
61663111@qq.com

Abstract. This paper studies the discourse analysis of Chinese cultural translation based on crawler algorithm. In recent years, the research of cultural translation has gradually become one of the research hotspots and difficulties in the field of translation. In order to improve the efficiency of discourse analysis of cultural translation and filter out most foreign translations with similar structure, an intelligent crawling algorithm is proposed. This paper first defines the concept of cultural translation, and then systematically combs the research of cultural translation. According to the location of nodes, the depth of DOM tree and the number of nodes with the same depth, the weight is assigned to each node, and then the similarity of cultural translation is calculated according to the given formula; On the basis of similarity, using the idea of aggregation hierarchical clustering, the cultural translation with similar structure is clustered into a group, leaving only one cultural translation in each group, so as to remove most cultural translation with similar structure. The experimental results show that the intelligent crawling algorithm can effectively reduce the cultural translation with similar structure and improve the inspection efficiency of vulnerability inspection system. On the basis of relevant data statistics, this paper summarizes the context, trend and existing problems of domestic research, and finally analyzes and prospects its current situation and work.

Keywords: Crawler algorithm · Cultural translation · Similarity

1 Introduction

Cross cultural communication and translation involve many important aspects, such as diplomacy, national image, cultural output and so on. Foreigners' misunderstanding of China's image caused by the wrong translation of Chinese Dragon into Western evil beast dragon has to make us deeply realize that cross-cultural communication and translation are indispensable. If cross-cultural communication and translation are not done well, we will be misunderstood, our culture will be plundered, and ill intentioned people will be given the opportunity to tamper with our national image [1]. In the field of cross-cultural communication, the practice of our ancestors is very brilliant. China's top strategy one belt, one road jointly built by China and other countries, has to thank Zhang Qian, who has opened up the Silk Road and the foresight of Emperor Han Wu. The silk road

M. A. Jan and F. Khan (Eds.): BigIoT-EDU 2022, LNICST 465, pp. 456–466, 2023.
https://doi.org/10.1007/978-3-031-23950-2_50

is not only the exchange of products, but also the dissemination of culture. China's four great inventions, advanced handicraft technology, advanced ideas and practical technology are disseminated to the world through this channel. The intersection with the Silk Road reflects Zheng He's feat of going to the West. Facing the open mind of the sea, the desire to share peace, the peaceful diplomatic policy of good neighborliness, leaving a civilized, friendly, generous and inclusive image of a big country to foreign countries, and spreading our most advanced technology and splendid culture are the cultural diplomacy that made the Ming Dynasty famous. There is no doubt about the great achievements of cross-cultural communication in ancient China, which makes foreign friends' impression of China still stay in the past. On the other hand, it also reveals the shortcomings of cross-cultural communication in today's China.

Our country's progress in military, economy, science and technology is indeed commendable, but the cultural and cross-cultural communication that affects the national accumulation is lacking. Our film market has become commercialized, bad films are popular, no acting skills, no connotation, let alone spreading culture. Chinese people still scold, not to mention foreigners. Plagiarism scandals are common in our academic circles. Cross cultural communication has a long way to go. In addition to learning from our ancestors' spirit of openness, inclusiveness and bold going out. In reality, we can also learn from European and American blockbusters, TV dramas, Japanese animation and Korean variety shows that are very popular with Chinese people, and integrate our own culture into spiritual production [2]. The popularity of these foreign cultures in China can be said to be cultural infiltration or even cultural invasion to a certain extent. If cross-cultural communication does not attract attention, let alone spread culture to the world, the status of Chinese culture in the country will also be threatened.

In view of the above situation, this paper proposes an intelligent crawling algorithm based on URL de duplication and aggregation hierarchical clustering based on cultural translation similarity for Chinese cultural translation discourse analysis, which can effectively remove repeated URLs and a large number of cultural translations with similar structure, minimize the test target of inspection system and improve the detection efficiency.

2 Related Technologies

2.1 Crawler Algorithm

Friends who have studied website design know that websites are usually designed in layers. The top-level domain name is followed by sub domain names, and there are sub domain names under sub domain names. At the same time, each sub domain name may have multiple domain names at the same level, and there may be mutual links between URLs, so as to form a complex network. When a website has many URLs, we must design the URL well, otherwise it will be very chaotic in the later understanding, maintenance or development process. After understanding the above structural design of cultural translation, the depth first algorithm in web crawler is formally introduced, as shown in Fig. 1 below.

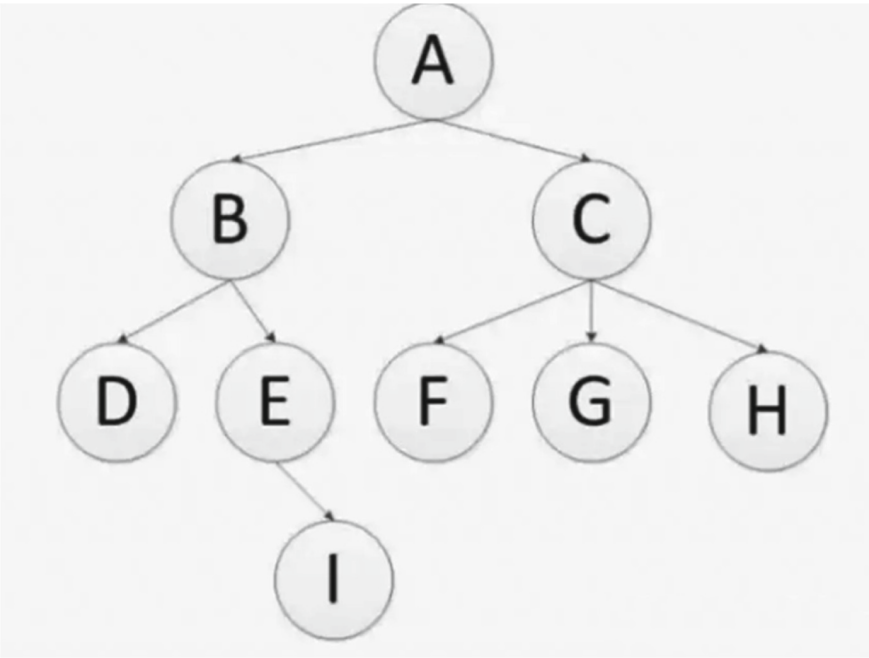

Fig. 1. Depth first algorithm tree structure in crawler

The above figure is a tree structure. Through the traversal of this tree, we can compare the external translation of grasping culture to deepen the understanding of reptile strategy [3]. The main idea of the depth first algorithm is to first start with the top-level domain name a, and then extract two links B and C. after the capture of link B is completed, the next link to be captured is d or E, not to grab link C immediately after the capture of link B. After capturing link D, we find that all URLs in link d have been accessed. Before that, we have established a list of accessed URLs, which is specially used to store the accessed URLs. When link D is completely crawled, link e will be crawled next. After the crawling of link e is completed, you will not climb link c, but continue to climb link I in depth. The principle is that the link will climb down step by step. As long as there is a sub link under the link and the sub link has not been visited, this is the main idea of the depth first algorithm. Depth first algorithm is to let the crawler go down step by step. After the crawl is completed, it will step back step by step and give priority to depth. After understanding the depth first algorithm and looking at the figure above, we can get that the order of crawler crawling links presented by the binary tree is: A, B, D, e, I, C, F, G and H (here, it is assumed that the links on the left will be crawled first). In fact, we often use this algorithm in the process of web crawler. In fact, our commonly used scrapy crawler framework is also implemented by this algorithm by default [4]. Through the above understanding, we can think that the depth first algorithm is essentially implemented by recursion. Figure 2 shows the code implementation process of the depth first algorithm.

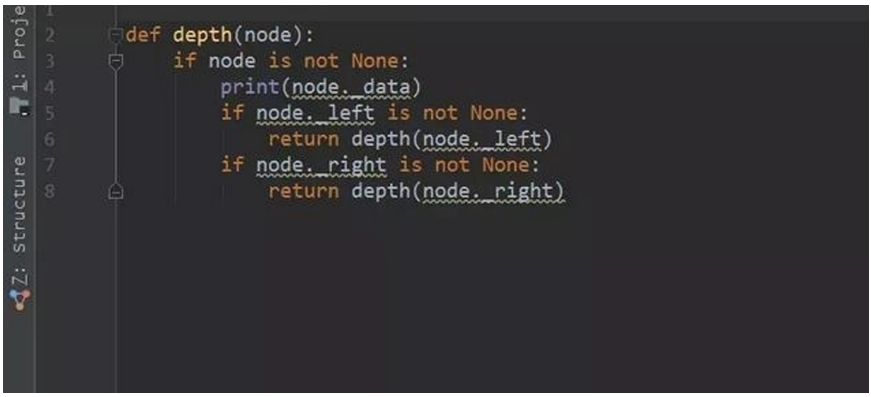

Fig. 2. Code implementation process of depth first algorithm

The depth first process is actually implemented in a recursive way. Looking at the code in the figure above, first define a function to implement the depth first process, and then pass in the node parameters [5]. If the node is not empty, print it. It can be compared with the top point a of the tree. After printing the node, check whether there are left node (link b) and right node (link c). If the left node is not empty, return it, call the depth first function itself again for recursion, and get new left node (link d) and right node (link E), and so on, It will not stop until all nodes are traversed or meet the established conditions.

2.2 Analysis of Cultural Translation

In a narrow sense, cultural translation is an activity to spread national cultural forms and products to target language countries through the medium of target language. In a broad sense, because language and culture are inseparable, all translation activities are essentially social and cultural communication activities. Therefore, all translation activities can be regarded as cultural translation activities (Zhong Weihe, Feng man 2014:57). The going out of Chinese culture requires not only the foreignization translation of Chinese translators, but also the domestication translation of foreign sinologists (Wang Qinghua 2015:100). In short, in the current situation of China's "going global", foreign translation is an export-oriented communication activity that takes the world as the development space and exchange platform for China's "going global" [6], translates China and Chinese culture into the target language and carries out cross-border, cross language and cross-cultural communication to the target language countries.

In order to facilitate analysis and comparison, based on relevant materials and data, this paper divides the existing situation of cultural translation research into three main aspects: scientific research institutions, talent training and academic achievements, in order to have a panoramic overview of cultural translation research in China.

In short, cultural diplomacy is a diplomatic activity in which independent sovereign states use cultural means to achieve specific purposes. It is an international activity. Its goal is to carry out peaceful and warm persuasion with the help of humanistic care and emotional investment, in order to establish a lasting and far-reaching international

friendship with foreign people and government officials, so as to obtain their sympathy, understanding and support [7]. This requires us to face up to and respect national differences, language differences and cultural differences, and pay more attention to the ways and skills of persuasion. In order to attract the foreign public and make it easy for them to understand what we think, it is wise to adopt the strategy of peaceful and flexible publicity of cultural truth. Our cultural diplomacy should maintain cultural security; Our cultural diplomacy should help to establish a positive national image all over the world; Our cultural diplomacy should be neither humble nor arrogant. All these constitute the spirit of cultural diplomacy.

3 Discourse Analysis of Cultural Translation

As culture going out is a national strategy advocated by the Chinese government and an important way to enhance China's overseas cultural soft power and shape the image of China's cultural power, domestic scholars have focused on the new academic field of cultural external communication and published many research results. Looking at the research articles on translation communication at home and abroad since the new century, it can be seen that scholars' research attention is shifting from the study of translation ontology to the cross-border study of translation and communication. In the face of so many research literatures, four different approaches can be explored, namely: 1) research on translation strategies; 2) Research on the quality of translated works; 3) Research on external communication channels: 4) research on the communication effect of translation. This paper discusses the gains and losses of each research path, so as to correct the deviation, promote the academic achievements, better guide China's foreign translation practice and communication, and improve the communication effect.

3.1 Research on Foreign Translation Strategies

Huang Youyi believes that foreign translation should be close to the reality of China's development, the needs of foreign audiences for Chinese information and the thinking habits of foreign audiences; Wang Ning believes that the translation style advocated by the English world is "domestication". Whether it is literal or not, if it cannot be understood by ordinary readers, it will be ignored by the book market; Zhang Jian believes that in foreign translation, the translator should make necessary adaptations to the original text according to the specific context of the reader, and the adaptation strategies include addition and supplement, interpretation and adaptation and reasonable deletion; Xie Tianzhen believes that due to the phenomenon of "time difference" and "language difference" in cultural exchanges between China and the west, translators should adopt domestication strategy or abridge the original work to varying degrees to meet the readers' reading acceptance context: LV Shisheng believes that Lin Yutang's unique cross-cultural interpretation of a dream of red Mansions enlightens us that Chinese cultural translation must be adjusted and flexible, To reduce the resistance to entering the Western cultural framework. Scholars have brought the reader's cultural acceptance context into the research field of vision, realized that foreign translation is the collision of two languages and cultures, and need to take the reader's reading habits and cultural acceptance context into consideration in foreign translation. They emphasize that the reader's

understanding and acceptance of the translation should be taken as the focus in foreign translation, which highlights that translation is a cross language, cross-cultural Cross social communication behavior. If the target language readers are difficult to understand the semantics of the translation, or do not recognize the ideological content of the translation, the translation will not achieve the purpose of cross-cultural communication. From this point of view, focusing on the target language readers and strengthening the research on the target language readers' reading expectations, cultural context and the mentality of accepting foreign cultures will help to improve the quality and communication effect of foreign translation.

3.2 Research on the Quality of Translated Works

Taking the translation of classics as an example, Wang Xiaonong analyzes the shortcomings of translation and editing in English translation, and discusses how to improve the quality of translation, editing and publishing; Ren Dongsheng investigated the phenomenon of incomplete correspondence between the two language texts in the typesetting of the English translation of Jin Ping Mei, and put forward some suggestions on bilingual translation, editing and publishing from the perspective of national translation practice. From the discussion of these two scholars, the translation quality of foreign translation works involves many factors, and the focus has shifted to the factors of editing and publishing outside the text. The author believes that the acceptability, comprehensibility and accuracy of the translated works will affect the communication and acceptance effect of the translated works among readers. Sometimes, the translator's Chinese-English bilingual skills are profound, and the translated content is faithful and smooth, which can stimulate the readers' reading desire. However, in the printing and publishing process, if the editing and typesetting of the translation are poor, the binding is not good, or the cover picture copy violates the readers' cultural acceptance taboos, these will directly reduce the communication effect of the translated works among the target language readers, Serious will also be resisted by readers. Therefore, the quality of translated works is determined by two links: 1) the quality of the original translation; 2) the layout and binding design in the process of printing and publishing. The first link requires the translator to have good bilingual control ability and be familiar with bilingual cross-cultural knowledge, so as to better ensure the quality of translation. Sometimes, on the premise of correctly conveying the semantics of the original, it is necessary to make appropriate cross-cultural obstacles to the original, so as to make the translation conform to the reading acceptance habits and cultural context of the target language readers; The second link concerns whether the translated works can capture the readers' eyes and enhance the attention of readers when readers quickly scan a wide range of books, which is one of the key factors for the translated works to be selected and read by readers. If the cover design is novel, beautifully bound and in line with the readers' aesthetic and cultural context, it will help the dissemination of the translated works. On the contrary, it will hinder the dissemination and acceptance of the translated works. In other words, the quality of the translated works is determined by the translation quality of the translation and the design of the translation typesetting.

3.3 Research on External Communication Channels

Zhang Kun believes that the multi-channel characteristics of the external communication of the national image can not be ignored. The bridge crowd, organizations and media can play the channel function of external communication; Zhang Menghan analyzed that foreign young netizens like to use new media platforms, and believes that external communication should pay attention to short video channels and mobile communication, so as to comply with the trend of integrated communication; Xu Jun believes that in addition to book translation and publishing, we should also integrate various media such as digital publishing, film and television, and the Internet, and give full play to the role of new media and integrated media in foreign cultural communication. From these representative results, scholars believe that the external communication channels are not limited to traditional newspapers, periodicals and radio stations. They need to explore the integration of new and old media and multi-channel simultaneous interpreting so as to achieve full coverage of readers' and audience's information Qu Daojin's coverage and increase the breadth and depth of readers' access to relevant information. Especially at present, with the popularity of mobile Internet in the world, readers or audiences can log in to the Internet anytime and anywhere and retrieve a large amount of relevant information. The Internet and mobile clients are becoming the main channels for people to obtain information. Therefore, exploring the communication role of new media in external communication is becoming the focus of scholars today.

At present, the media communication in western countries has changed from traditional paper media to network media and digital media. Therefore, China's external communication should conform to the information society, the fast speed of information communication, the readers' pursuit of information timeliness and the information publisher Characteristics of interaction with readers. This requires that the communication of cultural translation works should turn to the external communication situation dominated by network media and supplemented by paper publishing and communication. In addition, the foreign communication channels should not ignore the localized communication of overseas Chinese associations, Chinese overseas culture and Confucius Institutes. They are familiar with the customs, cultural taboos and communication skills of foreign countries. To sum up, the external communication channels of China's cultural translation works can include traditional paper media communication, network media communication, overseas Chinese community communication, China's overseas cultural center communication and overseas Confucius Institute communication. These five external communication channels can form a situation of multi-channel collaborative communication dominated by network media communication, supplemented by other channels, This can enhance the external dissemination of Chinese culture.

3.4 Study on the Communication Effect of Translation

There are few studies on the communication effect of foreign translation works by domestic scholars. The following are three representative research results: Li Ning believes that the circulation and retention of Chinese classics translation is too small and the attention is not high, which affects the communication effect of the translation among American readers; Bao Xiaoying analyzes all links in translation communication and believes that

all links in communication are an organism that jointly affect the communication effect of the translation; Liu Yameng and Zhu Chunshen believe that the translation needs influential book reviews to guide overseas readers' reading choices, which will help to improve the communication effect of the translation in the target country. The above-mentioned scholars' research on the communication effect focuses on various factors affecting the communication effect, and the analysis is justified. However, the communication effect of the translated works is an abstract concept [8]. To what extent can the translated works be considered to have the communication effect when they are spread and accepted abroad? At present, there is no quantitative index to evaluate the communication effect. Generally, the academic circles believe that if the translated works are spread to the overseas target country, even if they enter the country's libraries and bookstores, they will achieve the purpose of communication. In fact, this understanding is wrong. If the translated works are published and distributed in the target country, it can only show that the translated works are likely to be read and accepted by the target language readers, and it can not ensure that the translated works will be read and understood by the readers, Even if the reader has read and understood the content of the translated works, it does not mean that the reader has accepted the content of the translated works. The ultimate purpose of our foreign translation, introduction and communication is to make the target language readers understand and accept the ideas, views, value orientation and Chinese culture of the translated works. Therefore, the communication effect of foreign translation and introduction can be divided into two levels: primary communication effect and advanced communication effect. The level of primary communication effect is that the translated works are read and understood by the target language readers. Whether the readers agree with the ideological content and values of the translated works is not within the scope of primary communication effect; The advanced communication effect is that the translated works are read and understood by the target language readers, and at the same time, the readers can recognize and accept the ideological content, values or Chinese culture carried by the translated works. At present, foreign translators and foreign translation publishing departments should correct the cognitive errors of foreign translation effect, improve the reading rate and understanding of target language readers of foreign translation works, and make foreign translation works widely spread among readers [9]. On this basis, let the translated works be read by the target language readers, recognize and accept the ideological content or values of the translated works, which is the ultimate goal of the study of external communication effect. Therefore, foreign translation and communication should focus on the above two levels of research in order to truly improve the effect of foreign communication.

4 Discourse Analysis of Chinese Culture Translation Based on Crawler Algorithm

The translation of Chinese books into foreign languages is a controversial topic. If anyone needs someone to translate, then Chinese people have no qualification and right to engage in Chinese foreign translation. It's possible that Chinese people's English level is lower than that of native English speakers, but who can guarantee that English scholars' Chinese understanding ability will be better than that of Chinese scholars?

Professor Xu Yuanchong proved that the English translation of classics is first-class in China with his rich practice [10]. At the same time, he also pointed out the limitations of foreign translators' understanding.

With the development of web applications, crawler technology is also improving. From the early general crawler to the current theme crawler and deep crawler, it meets the different needs of different users. The crawling process of the general crawler is relatively simple. Generally, the crawler starts from the initial URL, takes out the new URL from the 1-L list and starts crawling, and cycles in turn until the list to be crawled is empty or meets the set stop conditions. This method is simple and commonly used, but this kind of crawler only crawls the URL, which can not meet the additional needs of users.

Compared with the URL of the general psychological element, the deep crawler not only extracts the URL, but also parses it to remove the duplicate of the URL, which improves the crawling efficiency and avoids falling into a dead cycle. On the basis of deep crawler, this paper proposes an intelligent crawler to crawl all relevant cultural translations, and then use the intelligent crawling algorithm to crawl the website [11]. Finally, no duplicate URL is obtained, and the cultural translation corresponding to the URL pairs the URL to discard the duplicate URL. Next, the page similarity formula is used to calculate the similarity value of the pages corresponding to the two URLs in turn. Specifically, the page is parsed into a DOM tree [12]. Root zhe t allocates the depth of each section and the number of nodes with the same depth from he, allocates the weight to each node, and then calculates the similarity of cultural translation according to the given formula. Finally, based on the similarity, the cultural translation with similar structure is clustered into a group by using the idea of aggregate hierarchical clustering, and only the representative URL is selected for subsequent testing.

The calculation process of the intelligent crawling algorithm proposed in this paper is divided into three stages, as shown in Fig. 3. In the first stage, the URL needs to be de duplicated; In the second stage, cultural translation is analyzed and the similarity of cultural translation is calculated [13]; In the third stage, the cultural translations whose similarity meets the set threshold are clustered, and a URL is selected from each category as a representative for subsequent detection. The whole calculation process is called intelligent crawling algorithm.

Recognition conditions in international academic circles The development of translation theory and cultural intervention in translation have strongly criticized the view of some scholars at home and abroad, represented by British Sinologist Graham [14], that Chinese-English translation can only be "translated into" by English translators but not "translated out" by Chinese scholars It is emphasized that Chinese-English translation is not the patent of foreigners, and Chinese scholars and translators should boldly undertake this work.

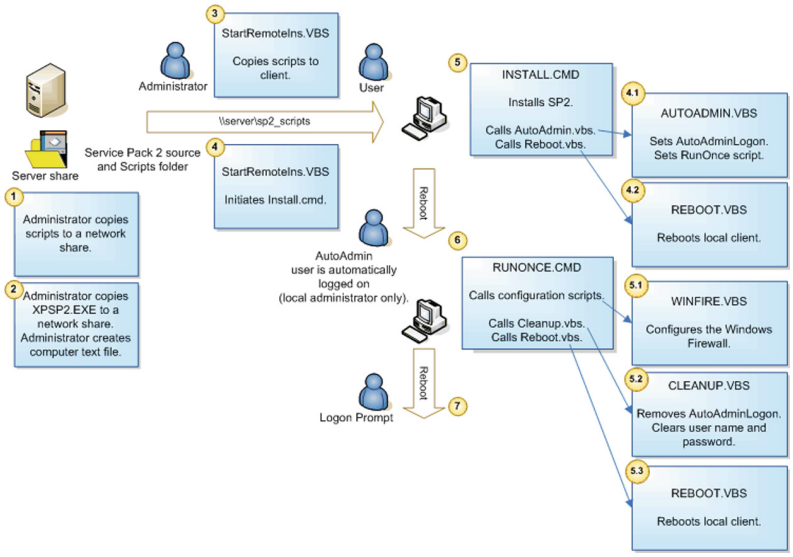

Fig. 3. Calculation process of intelligent crawling algorithm

5 Conclusion

In this paper, an intelligent crawling algorithm is proposed. In the crawling process, the Rabin fingerprint algorithm is used to remove the duplicate of the URL, which has fast retrieval speed and high efficiency, and avoids the crawler from wasting unnecessary time or even falling into a dead cycle; Before calculating the cultural translation similarity, the cultural translation is preprocessed, the DOM tree is constructed, and the text is uniformly represented by text, only the cultural translation label is retained, which simplifies the cultural translation, and the structural similarity of cultural translation is calculated by evenly distributing the weight, so as to make the calculation of cultural translation similarity more efficient; The aggregated hierarchical clustering algorithm is used to cluster the similar cultural translations into a group, which improves the efficiency of the inspection system. This paper solves the problem of a large number of similar cultural translation structures, and improves the efficiency of safety inspection. Moreover, this method has the advantages of simple idea, easy implementation, strong universality, high recognition rate of cultural translation with similar structures, and improves the accuracy of reptiles.

References

1. Wang, C., Pedrycz, W., Li, Z., et al.: Residual-driven fuzzy C-means clustering for image. Segmentation. **8**(4), 14 (2021)
2. Xuelei, W.: Advances in discourse analysis of translation and interpreting: linking linguistic approaches with socio-cultural interpretation. J. Commun. (2021)
3. Shaheen, F., Ali, G., Zahra, K.: Translating feminist identities: a critical discourse analysis of urdu translation of brown's work 'The Dancing Girls of Lahore' (2021)

4. Li, T., Hu, K.: Corpus-based translation studies and political discourse analysis (2021)
5. Li, T., Hu, K.: Reappraising self and others: A corpus-based study of chinese political discourse in english translation (2021)
6. A study on translation of etiquette culture with reference to two English versions of Hong Lou Meng. **28**(3), 8 (2021)
7. Ghader, H.: An empirical analysis of phrase-based and neural machine translation (2021)
8. Zhang, X.: A study of cultural context in Chinese-English translation. Region – Educ. Res. Rev. **3**(2), 11–14 (2021)
9. Bellassen, J., Wang, C.: Representation of Chinese language and culture through analysis of media coverage of linguistic education in France. Working Papers (2021)
10. Li, H., Xu, L., Fan, Y.: Cultural distance and customer orientation strategy of chinese service MNEs under the belt and road initiative. Cross Cult. Strat. Manag. (2021)
11. A brief analysis on the chinese translation of harry potter series: Gains and losses. **20**, 3 (2021)
12. Yetkiner, N.K.: A critical discourse analysis approach to othering: depiction of the syrian refugee experience in Turkish children's literature. [Sic] - A J. Literat. Cult. Liter. Transl. **2**, 11 (2021)
13. Alahyane, L.M.: Communicative translation & cultural globalization reading on Peter Newmark's Perception (2021)
14. Zhang, J.: Re-dissecting Ang Lee's crouching tiger, hidden dragon from the perspectives of cognition, translation and reconfiguration of culture. Compar. Liter. East West **1**, 1–20 (2021)