Development of Computer Intelligent Proofreading System from the Perspective of Medical English Translation Application

Yan Zhang(✉)

Xi'an Medical University, Xi'an 710021, China
yiyi800510@163.com

Abstract. With the increasing improvement of China's economic status and the increase of foreign exchanges of traditional Chinese medicine, the importance of medical translation has gradually increased. The English translation proofreading system based on phrase and syntax attaches importance to the accuracy calibration of phrase and syntax, fails to solve the problem of poor contextual coherence of English translation, Based on the development and research of computer intelligent proofreading system from the perspective of medical English translation application, this paper provides a scientific teaching idea for English practical teaching, combines translation theory with TCM translation practice, and better guides students' translation practice.

Keywords: Medical English translation · Computer · Intelligent proofreading system

1 Introduction

Since the reform and opening up, China's economy has developed greatly, its international influence has increased significantly, international cooperation and exchanges have become more and more extensive, and medical academic conferences, experience exchanges and medical reports have also increased rapidly. As an important medium of medical communication on the international stage, medical English plays an increasingly prominent role, and its text translation and research become more and more important. Standardized and accurate translation not only helps to learn foreign advanced medical knowledge and medical ideas, but also promotes the dissemination and exchange of medical technology and medical knowledge at home and abroad [1]. At the same time, it is also of great significance for medical English teaching and translation research.

Medical English is a very mixed subject, which covers the grammar of almost all countries in terms of vocabulary. When a thing is first discovered by people in a country, it is named after it in the language of that country. Therefore, both grammar and pronunciation are far beyond the scope of English. The name of this subject is "medical English", um In fact, I don't know who got the name. In short, we must first have a

© ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2023
Published by Springer Nature Switzerland AG 2023. All Rights Reserved
M. A. Jan and F. Khan (Eds.): BigIoT-EDU 2022, LNICST 466, pp. 211–222, 2023.
https://doi.org/10.1007/978-3-031-23947-2_23

concept, medical English, not English. Then, how long does it take for pure English majors to complete the translation work independently? Let me give you two examples: urgent and serious, and the difference between gout and stroke.

From the above two examples, you will find a problem. I write in Chinese, but it seems that you don't understand what it is in Chinese. Many students in my class are pure English majors. After graduation, it is difficult to learn because of the severe lack of science and engineering knowledge. The first to be eliminated in the class is British and American, which is a very strange phenomenon. Medicine is a dynamic subject, or a four-dimensional subject. Medical literature describes a dynamic process, so we must first know how to combine those things, and then we can accurately translate them.

Theoretically, it will take at least five to seven years to transfer from pure English to medical English. This is why few pure English learners turn to medical English translation. Generally speaking, medical literature is translated by medical learners themselves. Figure 1 below shows the translation of medical English body images.

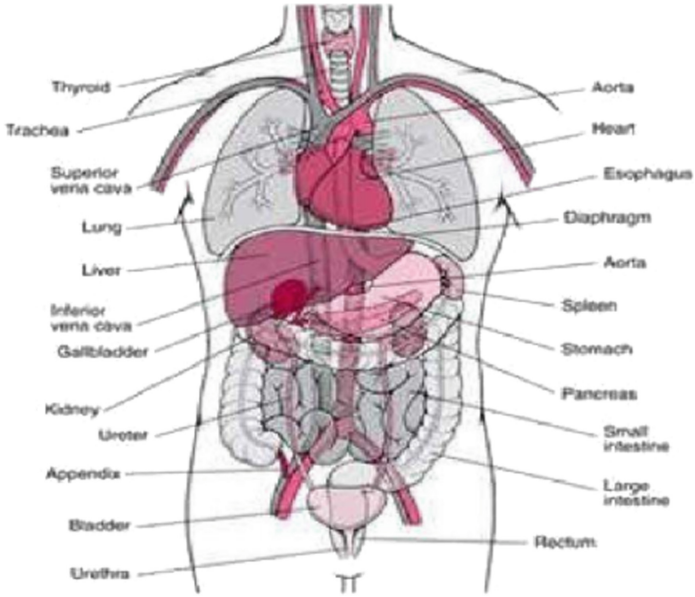

Fig. 1. Medical English body image translation

The development of Internet technology promotes the birth of a large number of English translation software. It is more convenient for people to obtain English translation results. According to its own algorithm settings, English translation software gives English translation results by searching the semantic vocabulary of the whole network. However, the reliability of these English translation results is poor and can not be used directly. A large number of manual proofreading is required in the later stage. There are two major defects in manual proofreading: the manual speed can not keep up with the proofreading demand of English translation results; With a large investment in manual proofreading, a large number of English translation computer intelligent proofreading

systems came into being. Feng Zhiwei adopts the English translation proofreading system based on phrase and syntax to proofread the English translation results, paying attention to the accuracy of phrase and syntax, but ignoring the coherence proofreading of context [2]; Li Yegang is an English translation and proofreading system integrating the largest noun phrase in human bilingual. Due to the lack of records of user behavior data, it is difficult for developers to optimize the system according to the user's use. Figure 2 below shows the framework of the English translation computer intelligent proofreading system.

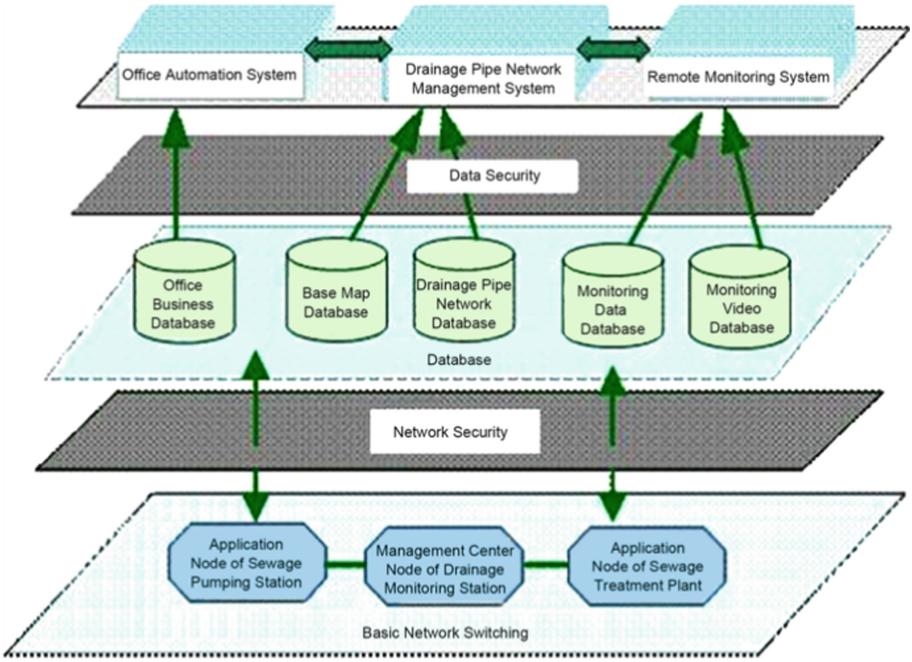

Fig. 2. English translation computer intelligent proofreading system

In view of the problems existing in the above system, based on the development and research of computer intelligent proofreading system from the perspective of medical English translation application, this paper introduces the functions and advantages of the system from two aspects of hardware and software, so as to realize the intelligent proofreading of English translation results.

2 Related Work

2.1 English Translation Computer Intelligent Proofreading Process Design

The process design of English translation computer intelligent proofreading based on phrase translation model mainly includes computer translation and intelligent proofreading. In terms of translation, after the system receives the input source text, it carries

out clause preprocessing analysis The sentence segmentation results shall be kept in the state of the whole sentence or phrase, which shall be used as the retrieval unit for input retrieval. Based on the database, the phrases similar to the searched phrases are matched, returned at the same time with the corresponding translation, and sorted according to the similarity. In order to effectively improve the retrieval efficiency and accuracy and facilitate the evaluation of similarity, the system indexes the corresponding corpora of source text and translation through auxiliary translation memory. In addition, the phrase translated by the translator may have high similarity with the phrase translated by other translators, or due to the influence of context, the phrase translated by the translator appears similar reproduction in the subsequent part of the text, so it is necessary to index and store the translated phrase.

For proofreading, load the source text translation text to be proofread, which can be read through the database. Text oriented, sentence segmentation first, and segment the source text and the translated Japanese standard text into phrases [3]. In translation, text segmentation follows the principle of segmenting into the shortest language segments, while in proofreading, the text needs to be segmented into phrases. The spaces in the English text are natural separators, so there is no need for word segmentation, just word segmentation for the Chinese clause text. There are various ways of Chinese word segmentation. This paper selects the Chinese word segmentation mode based on the maximum entropy model. After text word segmentation, nouns or phrases can be parsed for named entity recognition, and this information can be used as an anchor for English and Chinese word alignment. At the same time, part of speech is marked for the text to facilitate the identification of adjectives and numerals, in which numerals are the key information of text fragments to realize the alignment of English and Chinese words; Adjectives and adverbs are not only important carriers to enrich the emotional color of phrase fragments, but also important symbols of English and Chinese lexical alignment.

Complete the above in the text segment After processing, it becomes a phrase fragment set that has been divided into word boundaries, and words and phrases have been identified and marked by named entities, and then begins to propose phrase features to realize English Chinese vocabulary alignment. Moment language.

The feature extraction process is shown in Fig. 3.

For English and Chinese phrases, feature information is extracted according to the above process and described in digital form. In English Chinese bilingual vocabulary alignment, the distance between English and Chinese phrases is calculated by increasing the weight for each dimension, and the phrase distance is evaluated based on dynamic programming method, so as to take the English Chinese alignment model with the minimum phrase distance as the final alignment result [4]. To find the aligned phrases between the source text and the target file, and find the possible inconsistent expressions, that is, editorial errors or omissions in translation, so as to realize intelligent error detection.

The bottom-up proofreading module architecture is shown in Fig. 4.

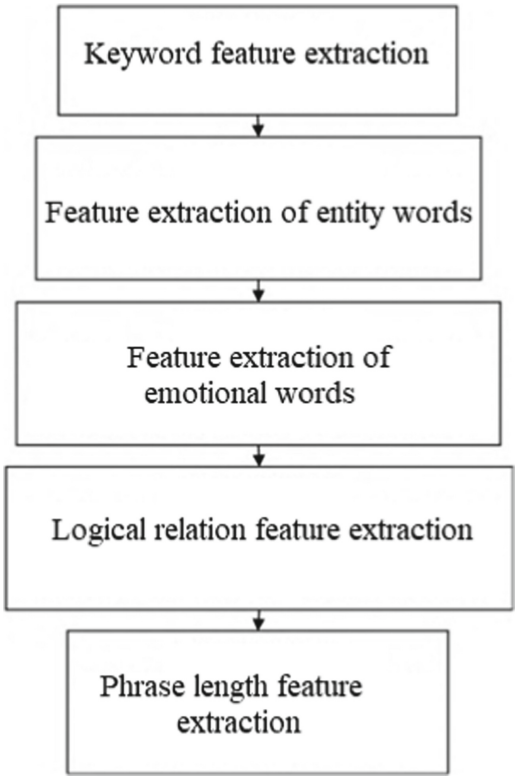

Fig. 3. Phrase feature information extraction process

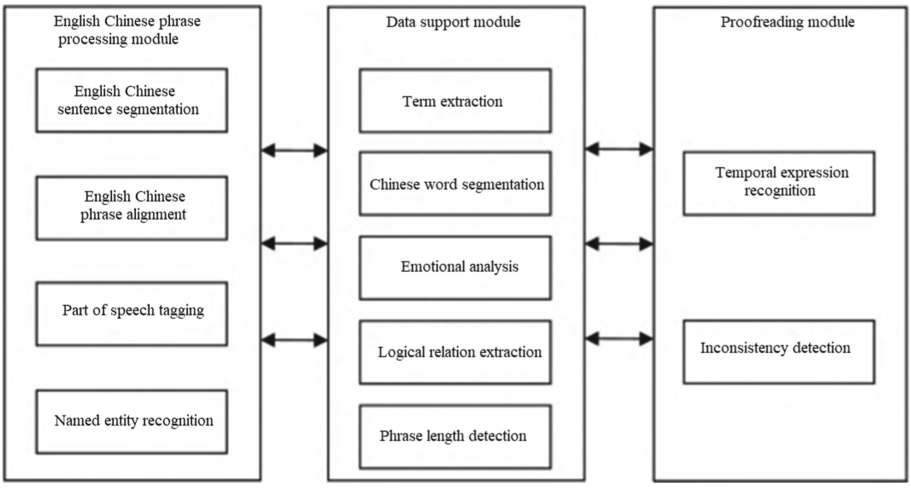

Fig. 4. Bottom up proofreading module architecture

2.2 Medical English Translation

With the improvement of science and technology and the gradual increase of international medical academic exchanges, medical English translation has become a very important subject. The demand for medical English translation is growing. To do a good job in medical English translation, we should not only master the knowledge of medical English, but also have good English level, translation skills and logical thinking ability. The following Fig. 4 shows the translation skills and logical thinking ability.

Firstly, medical English translation requires that the translation be consistent with the original content. The work of translation is to translate one language into another. The meaning of language is the same as that of the original language. Medical English translation is basically the same as translation. We must adhere to the principle of consistency and fluency. Therefore, all translations obtained from medical English translation should accurately and completely express the meaning of the original text without sentence confusion or ideology [5]. The situation is unclear.

Second, medical English translation requires correct grammar and smooth content. In order to better complete the translation of medical English, some grammatical errors should be avoided in the process of translation. Only when the grammar is completely correct can the meaning of the original text be expressed accurately and clearly. Therefore, medical English translators must have good knowledge of English grammar and be able to use it flexibly.

Third, medical English translation needs accurate translation and attaches importance to the translation of professional vocabulary. Medical English vocabulary has unique characteristics. It has strong professionalism. The meaning of many words is strictly limited by matching words and context. Therefore, medical English translation must be based on context and word collocation, combined with the article, give reasonable and accurate meaning, so that the whole translation can proceed smoothly. You can't think that every word has only one meaning. We should speculate according to the context and find the most appropriate meaning. Sometimes, in order to make the article easier to read, some parts of speech need to be converted. This puts forward higher requirements for the translator's medical English vocabulary, and we should pay attention to the understanding and application of vocabulary.

Fourth, in medical English translation, the attitude must be rigorous and pay attention to the sentence structure of translation. Medical English translation is a kind of scientific translation, which has high requirements for translation. In order to make the sentence structure rigorous, the content is logical, and the meaning of words is translated accurately [6]. Because most medical English articles emphasize the objectivity of content, mainly narrative and reasoning, many sentences in medical English articles will take on cumbersome and complex forms, such as passive sentences and inverted sentences. This requires the translator to pay attention to the sentence format, flexible conversion, and not forced translation according to the original sentence. Many medical articles involve personal information such as patient's name and doctor's name, and pay more attention to translation. We should not make mistakes to avoid misunderstanding.

The above are the four key points of how to do a good job in medical English translation, hoping to give some inspiration and help to medical translators. The following Fig. 5 shows the translation skills and logical thinking ability.

Fig. 5. Translation skills and logical thinking ability

3 Design of English Translation Computer Intelligent Proofreading System

3.1 Overall System Architecture Design

The overall architecture of the English translation computer intelligent proofreading system designed in this paper consists of work module, English translation module, English translation proofreading module, search module, user module and behavior log.

The work behavior data generated by the above five modules in the process of English translation intelligent proofreading can be recorded through the behavior log. The setting of behavior log provides a scientific basis for background engineers to view the footprint system in real time, and timely correct the problems existing in the working process of the research system, so as to improve the proofreading performance of the system.

English translation computer intelligent proofreading system is actually an English translation process. By translating English sentences and replacing the incorrect parts of the original translation results, it can realize intelligent proofreading and obtain the correct English translation results as much as possible [7]. The system searches for relevant translation information on the Internet according to the characteristics of the sentences to be proofread and stores it in the work module.

The function of the work module is the basis of intelligent proofreading of English translation. On this basis, it completes the intelligent proofreading of English translation. When the proofreading command is issued, the work module will receive the search link from the translation module. The English translation module will sort the translation results according to the similarity by analyzing the lexical features of the sentences to

be proofread, and finally select the most practical translation results. Users can view the translation results at the bottom of the order in the user module for reference.

3.2 Hardware Design

(1) Search module design. The extraction and analysis of lexical features in sentences is the main function of the search module.

When the search module receives the user input, it immediately launches vocabulary processing Work with feature search. The search module implements the basic meaning acquisition and subject content search of the words to be proofread by constructing the mapping thread [8].

(2) Behavior log. Behavior log is a record of all user behaviors in the system, which is presented in the form of data. When the user performs secondary proofreading, the behavior log records the user's footprint. If users proofread the same English translation many times, the system can intelligently increase the scope of Vocabulary Translation and search for more results that users may need, so as to improve the proofreading performance of the system and increase the accuracy of English translation computer intelligent proofreading.

3.3 Computer Intelligent Proofreading Method Based on Improved Phrase Translation Model

Therefore, the process of computer intelligent proofreading of English translation is actually the process of translating untranslated sentences, comparing and replacing the proofreading results with the initial translation results, so as to realize the intelligent proofreading of English translation [9]. The artificial neural network is introduced into the language model, and the continuous vector can replace the rough frequency with a relatively smooth probability. The model is established as shown in Fig. 6 below.

This paper defines h as the wrong English translation result and D as the correct English translation result. The transformation from h to D is the process of English translation. The English machine translation method based on the improved phrase translation model is as follows:

$$D = argmaxM\,(D|H) = argmaxM\,(H|D) \cdot M\,(D) \tag{1}$$

The accuracy of Vocabulary Translation in the results obtained by English machine translation methods needs to be improved, while computer intelligent English translation methods pay attention to the accuracy of vocabulary translation, that is [10], the accuracy of M(d) in Eq. (1). Therefore, on the basis of optimization formula (1) To realize computer intelligent proofreading on, the specific methods are as follows:

$$D' = argmaxM'(D|H)' \tag{2}$$

In order to facilitate the expression of computer intelligent proofreading method based on improved phrase translation model H is the word to be proofread, and the word to be proofread is represented by D. It is defined that there are p characters in H, represented by H8, which correspond to the vocabulary in the phrase translation model; At the same time, Q characters exist in D, represented by di.

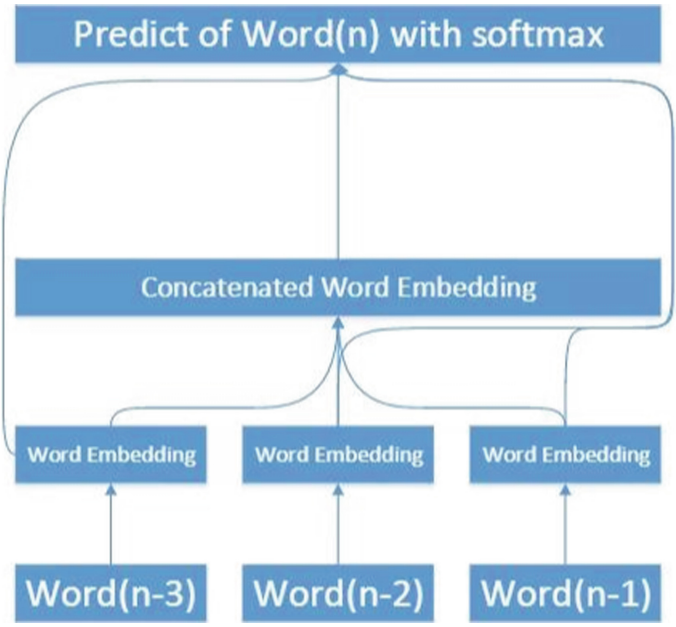

Fig. 6. Build language model

4 System Development

Medical English has experienced three main development stages: Old English, Middle English and modern English. Medical English has experienced several great changes in its long historical development, such as the Renaissance in Europe, the Norman conquest and so on. These events not only had a significant impact on politics, religion, education, justice and other fields in Britain, but also had an impact on the medical cause in Britain [11]. The direct result is that a large number of words and some structures from Latin, Greek and French have entered the medical field, become the main part of medical English, and gradually form a specific lexical structure system, specific grammatical rules and specific discourse structure. Medical English is characterized by accurate wording, objective expression, strict logic, concise writing, clear meaning, fixed meaning, and various strict application documents.

According to the classification of its subjects, medical English belongs to a member of Special English. The so-called special English refers to English related to a specific occupation, discipline or purpose. The difference from general English (also known as general English, mainly for the purpose of teaching language skills) is that special English has unique vocabulary, phrases, syntactic structure, habitual expression patterns and norms; Its purpose is to enable learners to specialize English knowledge and skills in a major or occupation based on general English.

The function of this module is to obtain the required HTML/XHTML web page code from the Internet according to the URL entered by the user, including a mobile phone login page in WML format and the web page code acquisition part of the CGI program

called by it. The code of mobile phone login page in WML format is shown in Fig. 7 below:

```
<?xml version="1.0"?>
<!DOCTYPE wml PUBLIC "-//WAPFORUM//DTD WML 1.1//EN"
"http://www.WAPFORUM.org/DTD/wml_1.1.xml">
<wml>
   <card title="Html2Wml" id="01">
      <do type="accept" label="转换">
         <go href="cgi-bin/wap.py?address=$(web_site)">
      </do>
      <p align="center"> Enter web site<br/>
         <input name="web_site" value="" title="web_site"/>
      </p>
      <p align="left">
         <small> Current versions:<br/>
                 html2wml&#x2122; v1.0<br/>
                 <br/>
                 <a href="about.wml">&#xA9; Copyright BTBU
Liujie</a>
         </small>
      </p>
   </card>
</wml>
```

Fig. 7. The Pseudocode for build language model

In page conversion, because the coding of Chinese characters is different from English, page conversion should consider the recognition, conversion and display of Chinese character coding [12].

The character set of XML document is the general character set of ISO10646. At present, this character set is consistent with Unicode. WML inherits the character set of XML document. WML pages do not require full Unicode encoding. Documents that do not use UTF-8 or utf-16 encoding must declare their encoding method in XML (UTF is the abbreviation of Unicode/UCS transformation format).

Unicode is a double byte character that can represent most characters in most languages used today, including Chinese characters that we are most concerned about. UTF-8 is an unequal amplitude encoding method. In UTF-8, ASCII characters remain unchanged and are not affected at all, but other Chinese characters need to be converted by programs. Moreover, a Chinese character needs three bytes to represent. It distinguishes the length of character encoding by setting the highest bit of byte. The first bit of single byte encoding is 0; The first bit of double byte encoding is 110, and the first bit of the second byte is 10; The first bit of the three byte code is 1110, and the first bit of the second and third bytes are 10. The coding principle and characteristics of UTF-8

make it easy to make two-way free conversion between UTF-8 and Unicode. Therefore, they are consistent in coding [13].

The encoding of Unicode is different from Big5 and GB2312 commonly used by us, and the support of some WAP micro browsers for bic5 and GB2312 character sets is not ideal. Therefore, Big5 and GB2312 should be converted to unicode encoding in some cases.

5 Conclusion

Medical English, as a kind of English for science and technology, has strong knowledge and professionalism. In the process of translation, we not only need to have a clear understanding and grasp of the source language, but also need to carry out more translation practice combined with medical professional knowledge. Under the guidance of the translation theory of text typology, analyzing and studying the translation of medical English can not only have a better understanding of the vocabulary and phrase characteristics of medical English, but also better analyze and understand medical English from the sentence level, so as to better promote the dissemination and exchange of medical knowledge, serve medical English teaching and translation research. Based on the development and research of computer intelligent proofreading system from the perspective of medical English translation application, this paper uses translation model to realize the intelligent proofreading of English translation.

Acknowledgements. 2021 Shaanxi Provincial Foreign Language Special Subject "Medical College and University English Curriculum Module Construction and Ability Training of Medical Students" (Project No.: 2021ND0640).

References

1. Upadhyay, U., Chen, Y., Hepp, T., Gatidis, S., Akata, Z.: Uncertainty-guided progressive GANs for medical image translation. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12903, pp. 614–624. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87199-4_58
2. Paavilainen, P., Akram, S.U., Kannala, J.: Bridging the gap between paired and unpaired medical image translation. arXiv e-prints (2021)
3. Chen, J., Wei, J., Li, R.: TarGAN: target-aware generative adversarial networks for multi-modality medical image translation. In: de Bruijne, M., et al. (eds.) MICCAI 2021. LNCS, vol. 12906, pp. 24–33. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87231-1_3
4. Larrahondo, B.F., Valencia, J.G., Martínez-Villalba, A.M.R.: Validation of the Self Stigma of Seeking Help (SSOSH) scale in a population of Colombian medical students. Revista Colombiana de Psiquiatría (Engl. Ed.) **50**, 82–91 (2021)
5. Tripathi, V.R., Kumar, J.H., Manish, P., et al.: Clinic, community, and in-between: the influence of space on real-time translation of medical expertise by frontline healthcare professionals in marginal tribal communities. J. Prof. Org. **8**(3), 3 (2021)
6. Tarek, M., Rahim, C.M., Nazrul, I.M., et al.: Translation, cross-cultural adaptation and validation of the English Lequesne Algofunctional index in to Bengali. Health Qual. Life Outcomes **18**, 1–9 (2021)

7. Jespersen, A.P., Lassen, A.J., Schjeldal, T.W.: Translation in the making: how older people engaged in a randomised controlled trial on lifestyle changes apply medical knowledge in their everyday lives. Palgrave Commun. **8**, 1–9 (2021)
8. Min, J.: Research on the application of computer intelligent proofreading system in college English teaching. J. Phys. Conf. Ser. **1915**(3), 032078 (2021)
9. Gao, J., Guo, Z.: Application of text proofreading system based on artificial intelligence. In: Atiquzzaman, M., Yen, N., Xu, Z. (eds.) BDCPS 2020. AISC, vol. 1303, pp. 722–727. Springer, Singapore (2021). https://doi.org/10.1007/978-981-33-4572-0_104
10. Liu, H.: Research on computer simulation big data intelligent collection and analysis system. J. Phys. Conf. Ser. **1802**(3), 032052 (2021)
11. Li, Y., Zhang, M., Chen, C.: A deep-learning intelligent system incorporating data augmentation for short-term voltage stability assessment of power systems. arXiv e-prints (2021)
12. Wu, W., Berestova, A., Lobuteva, A., et al.: An intelligent computer system for assessing student performance. Int. J. Emerg. Technol. Learn. (iJET) **16**(2), 31 (2021)
13. Sabri, Z.S., Li, Z.: Low-cost intelligent surveillance system based on fast CNN. PeerJ Comput. Sci. **7**(11), e402 (2021)