

Medical Literature Search Based on Database and Medical English Translation with Machine-Assisted Tool

Yanyu Chai¹, Yuhan Zhang², Yanxia Long^{2,*}

¹Graduate School of Translation and Interpretation, Beijing Foreign Studies University, Beijing, China

²School of Foreign Studies, Southern Medical University, Guangzhou, China

Email address:

sisulyx@163.com (Yanxia Long)

*Corresponding author

To cite this article:

Yanyu Chai, Yuhan Zhang, Yanxia Long. Medical Literature Search Based on Database and Medical English Translation with Machine-Assisted Tool. *International Journal of Applied Linguistics and Translation*. Vol. 9, No. 4, 2023, pp. 130-137.

doi: 10.11648/j.ijalt.20230904.14

Received: September 6, 2023; **Accepted:** November 8, 2023; **Published:** November 17, 2023

Abstract: In the context of the vast and diverse landscape of medical literature, the effective reading and translation of medical texts have emerged as pivotal aspects of medical English education and the initial step in facilitating medical research. Information retrieval, data compilation, and synthesis constitute integral components of this process. Furthermore, comprehending the intricacies and distinctions between Chinese and English, coupled with mastering fundamental translation techniques, represents a crucial challenge in the realm of medical English instruction. This study addresses the conundrums of swiftly identifying pertinent information and surmounting language barriers to produce precise and fluent translations. Building upon the foundation of the Themed Medical English Project, the research endeavors to offer various strategies for medical literature search, methods for refining search outcomes, and techniques for translating medical English. The objective is to enhance the quality and efficiency of medical literature comprehension and translation. Through this work, the research aims at empowering individuals engaged in medical research and English language learning to navigate the intricacies of this specialized field and to bridge the linguistic gaps that often impede the seamless exchange of medical knowledge. In conclusion, the research introduces two search query methods, including keyword and controlled vocabulary search, summarizes the characteristics of medical English and the resulting difficulties in translation process, and identifies two processes of machine-assisted human translation, including pre-translation adjustment and post-translation human editing.

Keywords: Medical Translation, Data Integration, Information Search, Machine Assistance, Artificial Adjustment

1. Introduction

Faced with mixed data and information online, students should select helpful information while remove useless message. At the same time, information integration of different categories is imperative to better outline the relevant aspects of one subject. Massive information online could be relevant to a disease to some extent, but may not be the most wanted for students. After searching for so long without targeted results, students could be overwhelmed. To increase students' efficiency in information searching and integration, proper methods should be used. According to Li, in the scientific treatises, there are mainly two ways, namely the key word index and the subject index. Key words are words,

phrases extracted from the title, abstract, or text of a document. These words or phrases can reveal the content of an article or a paragraph. It is an identification system for indexing entries. It is more flexible and has wider coverage than other index messages, and can reach the central content of the document more accurately. Another important advantage is that in commonly used online searches, keywords have good compatibility, and can significantly improve the precision ratio. In the subject index, subject words are different from natural language. There are a large number of synonyms, homographs, etc. in "natural language", which are different in concepts, languages, and things. There is often no one-to-one correspondence between objects. Sometimes, the same concept can be expressed with different words. If these "natural language" words are used as subject

words in the indexing or retrieval process, the purpose of one-to-one correspondence cannot be achieved, and it will bring difficulties to the search. To this end, it is necessary to select words from natural language vocabulary that can clearly express the theme of the document, have univocal meanings, and are used most frequently as subject words. [1] Based on previous research, this paper presents keyword searching and controlled vocabulary searching.

The second step of the themed project of ten major infectious diseases in human history is to translate medical English into Chinese with the help of machine translation. Medical terminology emphasizes subject terminology and industry terms in the medical field. Medical words are denotative languages, emphasizing the practicality and effectiveness of language, and pursuing semantic transparency, that is, certainty of reference and clarity of expression. In order to accurately express medical concepts, disseminate medical knowledge, and avoid confusion in medical terms, medical terminology must be strictly determined in terms of professionalism and accuracy. Medical English terminology has established a complete set of terminology standardization systems through word formation techniques such as incremental derivation, class transfer, synthesis, abbreviation, and reverse formation. [2] A comprehensive understanding of the types and word formation characteristics of medical vocabulary plays a vital role in understanding and translating medical English literature. [3] Chinese and English bear huge differences in vocabularies, grammar, sentence structure and discourse arrangement. Due to special features of medical English, several aspects apart from the differences mentioned above should be highlighted, including nominalization and passive voice. Machine translation includes a terminology management and memory system, vocabulary tools and search tools, as well as a search tool for typos. Machine-assisted translation mainly uses terminology management systems to correct and standardize terminology, ensure normalization of translation and accuracy of translation, and improve translation quality. [4] With the development of technology and expansion of corpus, it is improving in its accuracy and fluency, but there still exists drawbacks and mistakes. Analysis of these differences and translation methods to these difficulties would be presented in the paper to adjust the version of machine translation and enhance the professionalism and accuracy of translation.

2. Medical Literature Search

2.1. Medical Literature Search and Data Integration Before Translation

There are mainly three problems in searching and

collecting references: first, lack of pathways to finding medical literature due to deficiency in knowledge of search engines or medical database; second, lack of precision and efficiency in data integration due to the large amount and diversity of references; third, difficulties in comprehending and citing the references due to language difference and complexity of terminology in English or source language. Effective methods of literature search and information collection are needed.

Two basic search queries are listed below: Keyword and Controlled Vocabulary.

2.2. Keyword Searching

First, the Keyword searching refers to entering a search term that readers believe best describes the topic used in a reference, in which keywords could be the words the author selects, word in the title or the abstract or the word anywhere in the reference. While selecting keywords, consider all the different options, including synonyms and variations of the search term to make sure the retrieval of all the relevant records. [10]

Keyword is also known as non-indexed word. Different databases would use different subject terms to describe the meaning of keywords. For example, “heart attack” or “cardiogenic shock” would be used as “heart disorders” in the database PsycINFO and Myocardial Infarction in the MeSH term. Problem for keyword searching is that due to diversity and professionalism of medical terms and disease descriptions, if reader could not get the exact word that author uses, it would be difficult to find the article needed.

To optimize the keyword searching, all the relevant terms should be considered to guarantee the completeness of the searching result. All those terms should be added to the search with an OR relationship with terms that are already used. If relevant terms are not included, the search results would not be specific and interrelated enough, thus adds difficulties to reference reading and selection. One solution to that is to conduct a quick study and scan the top retrieved articles in search results sorted by relevance. In reading the top articles, try to look for other synonyms or relevant terms to add to the search strategy. Because the initial search aims at completeness of all the relevant articles, the core of quick study lies in finding potential relevant terms, rather than thoroughly study the whole article. [11]

Take COVID-19 as an example. The target of the retrieval is the psychology effects caused by COVID-19.

Enter keyword “COVID-19” and relevant term “psychology”, we could get the initial search results.

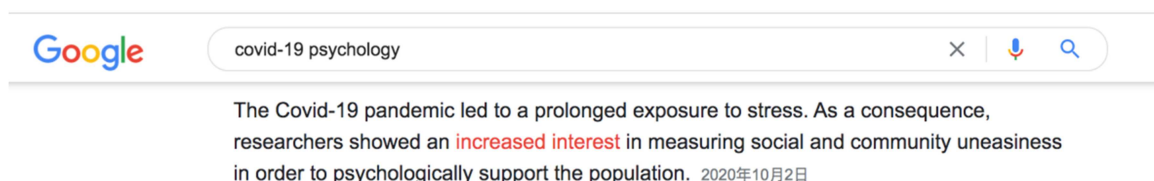


Figure 1. Initial search results of “COVID-19” and relevant term “psychology” [7].

What we can get from quick study related the search terms is that 1. the pandemic would lead to stress 2. major measurement is on social uneasiness.

Then, take stress as a searching example, and change the relevant terms to “stress” We would get results as follows.

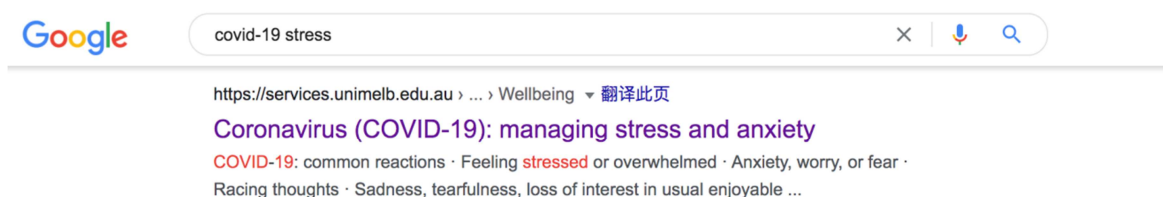


Figure 2. Search results of “COVID-19” and “stress” [8].

Now, we could get more specific results. More detailed information and psychological symptoms caused by COVID-19 could be found, including frustration, irritability, anger or restlessness. Additionally, by further digging into the article, strategies to cope with stress and anxiety could be learned.

When conducting keyword searching, we should not be too focused on one single keyword, but to analyze and scan the search results to get more relevant terms, thus to be more efficient and time-saving medical literature search.

2.3. Controlled Vocabulary Searching

Second, the Controlled Vocabulary searching is also known as the Indexed Vocabulary searching or Thesaurus searching. Controlled vocabularies are standardized, hierarchical lists designated to represent the specific subject concepts and conditions contained in a database. The hierarchical lists of controlled words benefit searching by allowing broad concepts to be narrowed down to more specific terms in that framework or allowing specific terms to be generalized to broader terms. For example, searching typhus in Medical Subject Headings (MeSH) of PubMed. MeSH terms are organized into a hierarchy called the MeSH tree. When search typhus in PubMed, there are three types of typhus: Typhus, Epidemic Louse-Borne, Typhus, Endemic Flea-Borne and Scrub Typhus.

Looking through their MeSH trees, we could generally identify the way they are grouped and their differences according to their classification. (PubMed - Searching Medical Literature: Basic PubMed Searches, 2020).

Scrub Typhus is caused by Rickettsiaceae infections. Rickettsiaceae infections could be subdivided into Rickettsia infections, which causes Typhus, Epidemic Louse-Borne and Typhus, Endemic Flea-Borne. And looking into Rickettsiaceae Infections, we can see it can be subdivided into another disease apart from Rickettsia Infections and Scrub Typhus, Pneumonia, Rickettsial.

Controlled vocabulary varies from database to database. Indexed databases generally provide their own thesaurus, e. g. PubMed: Medical Subject Headings (MeSH), PsycINFO: Thesaurus of Psychological Index Terms. Take MeSH as an example. MeSH terms are official words or phrases chose by experts or subject specialists to represent particular concepts. When labeling an article, indexers select terms only from the official MeSH list, which avoids variations and different spellings. By following such strict rules, searching could be efficient. In indexed databases, no matter what term an author may use, there will be one subject heading applied for that topic, and in MeSH, we only need to use one subject term to collect all indexed articles.

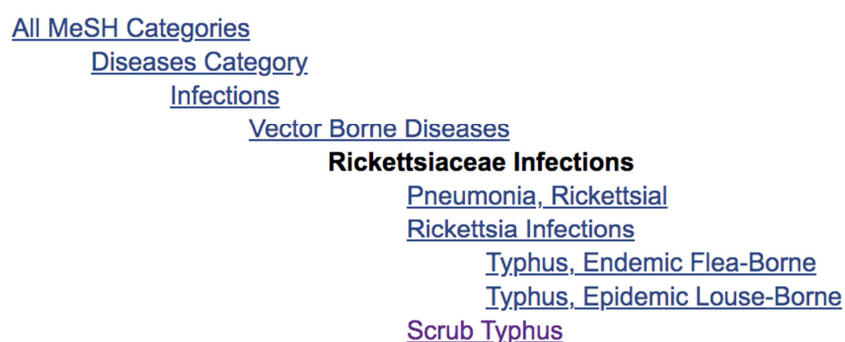


Figure 3. MeSH tree of Rickettsiaceae infections [6].

MeSH terms can also help us locate articles exactly focus on THE topic, rather than those just mentioning the topic. For example, searching for the word “cancer” would receive irrelevant but a majority of results, such as “people with cancer were excluded as the majority of malignancy are a contradiction for organ donation”. However, if search MeSH term Neoplasms, there would only be articles specifically

about cancer, which could increase efficiency and improve precaution.

2.4. Use of MeSH Search

1. Identifying additional aspects of the subject terms, thus we could narrow down search results. The additional aspects would include pathology, transmission, therapy etc.

It is important to know the pathology of the disease. Only by knowing the pathology of the disease, can it truly be understood and classified. For instance, Scrub Typhus is caused by Rickettsiaceae infections, while Typhus, Epidemic

Louse-Borne and Typhus, Endemic Flea-Borne are caused by Rickettsia Infections. Different pathology would affect prevention or treatment strategies.

The screenshot displays the PubMed Search Builder interface. At the top, there is a 'Send to:' dropdown and a 'PubMed Search Builder' header. The search criteria are entered in a text box: `("COVID-19/history"[Mesh] OR "COVID-19/mortality"[Mesh] OR "COVID-19/therapy"[Mesh])`. Below the text box are buttons for 'Add to search builder', 'AND', and 'Search PubMed'. To the right of the search box is a 'YouTube Tutorial' link. Below the search box is a 'Related information' section with links to 'PubMed', 'PubMed - Major Topic', 'Clinical Queries', and 'NLM MeSH Browser'. At the bottom is a 'Recent Activity' section with a 'Turn Off' and 'Clear' link, and a list of recent searches: 'COVID-19', 'COVID-19/history (0)', 'covid-19 (110)', and 'covid-19 vaccine (16)'. On the left side, there is a 'Full' dropdown and a 'COVID-19' section with a description of the disease. Below the description is a 'PubMed search builder options' section with a 'Subheadings:' link. The subheadings are listed in three columns, each with a checkbox: analysis, anatomy and histology, blood, cerebrospinal fluid, chemically induced, classification, complications, congenital, diagnosis, diagnostic imaging, diet therapy, drug therapy, economics, embryology, enzymology, epidemiology, ethnology, etiology, genetics, history (checked), immunology, legislation and jurisprudence, metabolism, microbiology, mortality (checked), organization and administration, parasitology, pathology, physiology, physiopathology, prevention and control, psychology, radiotherapy, rehabilitation, statistics and numerical data, surgery, therapy (checked), transmission, urine, veterinary, and virology. At the bottom left, there are two checkboxes: 'Restrict to MeSH Major Topic.' and 'Do not include MeSH terms found below this term in the MeSH hierarchy.'

Figure 4. Subheadings of COVID-19 [5].

Involving transmission is also important. For instance, COVID-19 spreads primarily when people are in close contact and one person inhales small droplets produced by an infected person coughing, sneezing, or talking. When the contaminated droplets fall to floors or surfaces they can, remain infectious if people touch contaminated surfaces and then their eyes, nose or mouth with unwashed hands. By searching the mode of transmission of the disease, we can precisely cut the transmission route, thus to reduce the infected and control the epidemic.

2. Once we decide all other relevant aspects, checking off all the terms we need, sending them to the PubMed search builder and combining the separate concepts or subheadings together with the AND commander. Click Search PubMed to complete the search. For example, COVID-19 Mortality, History and Therapy would help to find articles that are labeled with both words. [9]

2.5. Websites and Databases Recommended

To guarantee the credibility of source information, information should be retrieved in reliable websites and databases. Reliable sources are the foundation for data and references selection and integration. Major websites and databases are Google Scholar, Wikipedia, PubMed, PsycINFO, CINAHL, Embase.com, and ProQuest. Journals could also be another efficient source for references, including *Nature*, *JAMA (the Journal of the American Medical Association)*, *the Lancet* and *NEJM (the New England Journal of Medicine)*. Additional sources could be

government medical websites including CDC (www.cdc.gov), NHS (www.england.nhs.uk) and WHO (www.who.int). We could also get references from the source reference which are listed below the medical literature. It is also a convenient way to get references.

3. Medical Literature Translation

3.1. The Characteristics and Difficulties of Medical English

Medical English, because of its special characteristics, tends to cause difficulties for translators in the process of translation.

3.1.1. Grasp of Common Used Vocabulary

For one thing, the most commonly used words in medical English are simple words, such as do, get, take. But these words have the most abundant meaning and usage and functions. This also makes it difficult to grasp the exact meaning of the word in the process of translation.

3.1.2. Common English Words Are Borrowed

Second, medical English borrows some common English words. These words often have special medical meanings in medical English. For example, "complaint" means "expression of dissatisfaction" is well known, but in the medical context, it can also mean "the state of symptoms". And "solution" in common English means "a way of solving a problem or dealing with a difficult situation", but it also means "a liquid in which something is dissolved." If the

translator does not know these usages, it will easily lead to errors.

3.1.3. *Wide Use of Abbreviations*

Third, the wide use of abbreviations is also one of the most prominent characteristics of medical English translation. There are a large number of abbreviations in English vocabulary. For example, initialism such as “BP”(blood pressure), acronym such as “SARS”(Severe Acute Respiratory Syndrome) and clipped word such as cardiogram, which means electrocardiogram. Although the application of such words can reduce the complexity of English use, enhance its professionalism and reduce the difficulty of reading, it will increase the difficulty of translation. [15]

3.1.4. *Multiple Ways of Expression*

Fourth, in medical English vocabularies, a meaning can often be expressed in more than one ways including formal and colloquial expressions with different stylistic characteristics, which are suitable for different contexts and occasions. For example, " a disease of the heart " can be expressed simply as “heart disease”, or in technical terms as “cardiopathy”. Informal word “pee” and formal word “urine” can be both used to express the liquid excretory product. Different uses in various contexts require accurate discrimination by translators.

3.1.5. *Generation of New Vocabulary*

Fifth, with the continuous development of medicine, new disease names, drug names, therapeutic instruments and treatment methods appear at a very fast pace, which often leads to corresponding defects (对应缺损) in the process of medical English translation. When literal translation is difficult and the target readers cannot understand it, it is necessary to follow the rules of word formation of medical words and adopt translation methods of transliteration (aspirin - 阿司匹林), image translation (sigmoid flexure - 乙状结肠), morphological translation (X-ray - X线) and free translation (penicillin - 青霉素) to create equivalent Chinese words. [13]

3.1.6. *The Use of Many Root Words and Affixes*

Many medical English words are derivatives combined with prefixes, roots and suffixes. For example, “electrocardiogram” is the combination of prefix “electro-”(electricity), “-cardio-” (heart) and “-gram”(a thing that is written or drawn). Despite the convenience brought by affixes to comprehend a long and difficult word, it is difficult for medical English learners to memorize those words and affixes.

3.2. *Text Adjustment Before Machine Translation*

In order to improve the quality of machine translation output, the translator need to pre-process and modify the source text before translation.

First, spelling and punctuation checks are necessary to avoid translation errors caused by minor mistakes. Second, make manual changes to sentences whose modifying

relationship is not clear or may cause ambiguity. It can be achieved by replacing pronouns and synonyms and adjusting the order of words, etc. [12] In addition, long and complex sentences are common in medical professional English due to the need for tight structure and logical rigor. The translator needs to analyze the sentences carefully, understand the original text, and use sentence fragmentation to reduce the difficulty of syntactic analysis by the system and reduce uncertainties.

Adequate pre-editing is the basis for reducing errors in machine translation, and can also reduce the translator's workload in post-editing.

3.3. *Defects of Machine Translation*

Because of the complexity of natural language and translation and the limitations of the machine itself, machine translation still has some problems, such as structural ambiguity and semantic ambiguity. Structural ambiguity is manifested in messy sentences, grammatical errors, and incoherent meanings between sentences. Semantic ambiguity is manifested in improper wording, semantic errors, and even inconsistencies. [16] For example, in the sentence " people with SARS CoV must be isolated to limit the chances of medical personnel become infected ", the result of machine translation is “必须隔离患有 SARS-CoV 的人, 从而限制医务人员感染的可能性”. It can be seen that the terms are unprofessional and inaccurate. It should be changed to “必须隔离 SARS-CoV 患者以减少医务人员感染的几率” to be more consistent with the conciseness and professionalism of medical English, and to improve the accuracy of the wording.

3.4. *Artificial Post-Translation Adjustment*

As mentioned above, post-translation adjustments by the translator are needed to modify the results of machine translation. There are also several aspects of post-translation adjustments that require particular attention.

3.4.1. *Check Nouns And Terms*

First, the translator should check the translation of proper nouns and terms. Due to the specialized nature of medical English, the accuracy of medical terminology is particularly important, so this aspect needs to be checked more carefully by the translator. Some international academic institutions, for insurance, mostly have established Chinese translations, can not be changed at will. For example, International Society for Neurovascular Disease (ISNVD): 国际神经血管学会.

3.4.2. *Adjust the Order of Words*

Second, the translator should check the order of words. Depending on the expression habits of the target language, the correctness of the order is an important factor in improving the acceptability of the translation. Particular attention needs to be paid to the placement of subordinate clauses, gerunds, interjections, etc., which are often more difficult for machine translators to notice. For example, “Five to nine days after the symptoms have started, a rash typically

begins on the trunk and spreads to the extremities. This rash eventually spreads over most of the body, sparing the face, palms, and soles.”

Inappropriate translation: 出现症状之后的五到九天, 皮疹通常始于躯干, 并向四肢蔓延。皮疹最终几乎会蔓延全身脸部、手掌、脚底除外的各处。

Appropriate translation: 出现症状后的 5-9 天, 躯干开始出现皮疹, 并向四肢蔓延。皮疹最终几乎会蔓延全身, 但脸部、手掌和脚底除外。

The first version is not consistent with Chinese expression and there is also redundancy, which violates conciseness of professional English. After adjusting the word order, and take “但脸部、手掌和脚底除外” as a complement rather than an attribute, the second version fully demonstrates the meaning of English with clarity and simplicity.

3.4.3. Avoid Cultural or Ideological Conflict

Third, the translator should check whether there is cultural or ideological conflict. For example, when conducting an introduction to COVID-19, it is easy to see expressions on foreign websites that are politically offensive or culturally insulting. This is sometimes difficult for machine translation to recognize, so it requires the translator to be flexible and adaptable to avoid inappropriate translations.

3.5. Considerations in Medical English Translation

In addition to the scope for post-translation editorial changes described above, there are a number of considerations for medical English translation.

3.5.1. Treatment of Words

First, in the treatment of words, the general rule is to make sure words are precise, simple, concise and professional. Correct understanding and expression of the meaning of words is the prerequisite for the quality of the translation.

a. Handling of amphibious words

Amphibian medical vocabulary has different meanings in different professional majors, sentence structures and contexts, and requires a comprehensive judgement and choice based on sentence structure, context, medical jargon and customary Chinese expressions. For example, “The observation that maternal, more than paternal, asthma is a significant risk factor for the development of asthma and atopy in children might indicate that preventive events can even be traced back to preconceptional exposures of the mother.” was translated into that “观察发现相对于父亲, 母亲哮喘更是儿童发生哮喘和特应体质的危险因素, 这提示儿童哮喘预防要从母亲孕前开始”. In the common context, “preconceptional” means “Foresight, anticipation, prejudice”, but it means “before pregnancy” in this context. “Event” means “experience, activity”, but “Preventive events” in this sentence means “precaution”. In the process of translation, not every word in English can be found in Chinese with the same meaning and lexical identity, and it is often necessary to change the lexical identity in order to conform to the Chinese idiom.

b. Omit or add words

In addition, due to differences between Chinese and English, words or phrases should be omitted or added in translation. English is hypotactic, which requires complete sentence structure and word complement, so there are a large number of articles, prepositions, pronouns and conjunctions. However, Chinese is paratactic, which is characterized by chronicle style and focuses on its function and meaning. Therefore, in English, there are a large number of articles, prepositions, pronouns and conjunctions etc. All these words should be omitted. And other Chinese typical structural words or words to supplement meaning should be added to make the translation complete, accurate and smooth.

For example, “COVID-19 spreads primarily when people are in close contact and one person inhales small droplets produced by an infected person (symptomatic or not) coughing, sneezing, or talking.”

Inappropriate translation: COVID-19 主要在人与外界密切接触时传播, 吸入感染者(有症状或无症状)咳嗽、打喷嚏或说话时产生的飞沫。

Appropriate translation: COVID-19 主要在人与人密切接触时传播。当一个人吸入感染者(有症状或无症状)咳嗽、打喷嚏或说话时产生的飞沫时, 传播就发生了。

The first translation is a literal translation and follow the order of the sentence without adjustment. The sentence describes two situations that COVID-19 transmits. They should be separated into two independent sentences in Chinese rather than combine them together. “当一个人” and “传播就发生了” should be added to completely convey the meaning of the text.

3.5.2. Treatment of Sentences

Second, in the treatment of sentences, strong and clear expressions as well as correct tenses are vital. In order to make the translation close to the original text and to standardize the language, it is often necessary to convert the sentence components.

a. Use of passive voice

At the same time, in medical literature, the focus is on the object of the act, i. e. the disease and the patient, and rarely on the actor itself, i. e. the medical practitioner. Therefore, the passive voice is often used in medical English literature. The use of the passive voice is more objective in stating facts, avoiding the inclusion of personal subjective emotions, and provides the effect of prefacing information (信息前置) and highlighting meaning. [17] However, passive sentences are seldom used in Chinese, so we should pay attention to the necessary voice conversion in translation so as to make the translation more in line with the reading habits of readers.

b. Transformation of nominalization structure

Besides, Chinese medical literature often uses verbs to indicate the occurrence of an action, whereas medical English uses more nominalization structures, especially the extensive use of nouns derived from verbs, which can reduce the number of sentences or clauses while conveying a great deal of information. For example, The prevention of sudden cardiac death by long-term administration of antiarrhythmic drugs is based on sound physiological principles. (长期服用

- [15] Lu Y. & Wei R. (2020). Yi xue sheng yi xue yin yu fan yi neng li ti sheng de fang fa tan xi [Methods to improve medical English translation ability of medical students]. Ying yu guang chang, (33): 36-38.
- [16] Pan, Z. & Luo, H. & Yi, Y. (2017). Ji qi fan yi de kun jing qian jing he chu lu [The dilemma, prospect and outlet of machine translation]. Shan dong nong ye gong cheng xue yuan xue bao, 34 (09): 44-49.
- [17] Zhou, J. & Zhou Y. (2012). Ji suan ji fu zhu yi xue yin yu fan yi ji qi jiao xue [Computer-aided Medical English translation and its teaching]. Zhong guo yi xue jiao yu ji shu, 26 (05): 545-550.

Biography

Yanyu Chai (2000-), born in Zhejiang Province, China, graduate student in Graduate School of Translation and Interpretation, Beijing Foreign Studies University, Beijing, China. Having keen interest in research in English Translation and Interpretation, with current project of research on the effect of computer-assisted interpretation on the quality of simultaneous interpretation.

Yuhan Zhang (2000-), Born in Zhejiang Province, China. Bachelor of Medical English from Southern Medical University, Guangzhou, Guangdong Province, China, 2022. She currently works in projects in reading, translation and reverse writing of medical English literature by using corpus and artificial intelligence translation.

Yanxia Long (1979-), born in Chongqing, China, PhD in English Language and Literature, associate professor at Southern Medical University, Guangzhou, China.