

Postprint of Research on Terminology Standardization for Traditional Chinese Medicine Diagnosis and Treatment Knowledge Base

Authors: Ma Jie, Wu Wenling, Sun Hengyu, Wang Chunxiu

Date: 2023-04-01T16:02:44+00:00

Abstract

[Purpose/Significance] Traditional Chinese Medicine (TCM) terminology serves as the carrier of TCM knowledge and constitutes an important component of the TCM diagnosis and treatment knowledge base. Research on terminology standardization for the TCM diagnosis and treatment knowledge base is conducive to improving the quality of medical case data information in the knowledge base and promoting TCM culture. [Method/Process] This study determines the scope of terminology extraction and word selection for the TCM diagnosis and treatment knowledge base, constructs the structure of terminology standard files, proposes semantic control strategies for existing problems in terminology standardization, and illustrates the process of knowledge base terminology extraction and standard file establishment with specific examples. [Results/Conclusion] Terminology standardization work based on the TCM diagnosis and treatment knowledge base can enhance the utilization efficiency of medical cases in the knowledge base and advance the modernization and internationalization of Traditional Chinese Medicine.

Full Text

Preamble

Authors: Ma Jie^{1,2}, Wu Wenling¹, Sun Hengyu³, Wang Chunxiu¹

¹School of Management, Jilin University, Changchun 130022

²Information Resources Research Center, Jilin University, Changchun 130022

³The First Clinical Hospital of Jilin Academy of Traditional Chinese Medicine, Changchun 130022

Abstract: [Purpose/Significance] Traditional Chinese Medicine (TCM) terminology serves as the carrier of TCM knowledge and constitutes a vital component of TCM diagnosis and treatment knowledge bases. Research on terminol-

ogy standardization for TCM diagnosis and treatment knowledge bases helps improve the quality of medical case data information in these knowledge bases and promotes TCM culture. [Method/Process] This study determines the scope of term extraction and selection for TCM diagnosis and treatment knowledge bases, constructs the structure of terminology standard files, proposes word meaning control strategies for existing problems in terminology standardization, and illustrates the process of knowledge base term extraction and standard file establishment through specific examples. [Result/Conclusion] Terminology standardization based on TCM diagnosis and treatment knowledge bases can enhance the utilization efficiency of medical cases in these knowledge bases and advance the modernization and internationalization of TCM.

Keywords: TCM diagnosis and treatment knowledge base; TCM terminology; terminology standardization

Classification Number: G251.2

DOI: 10.13266/j.issn.0252-3116.2021.02.002

With the advent of the information age, internet and big data technologies have brought excellent opportunities for the development of TCM. In the issued “Strategic Planning Outline for the Development of Traditional Chinese Medicine,” the state explicitly instructs that by 2020, the task of enabling all citizens to access basic medical services should be achieved, vigorously promoting the integration of internet and TCM, and constructing a new model of online remote medical care [1]. A knowledge base is an artificial intelligence system that provides knowledge services [2] and has been widely applied in multiple fields. By constructing a TCM diagnosis and treatment knowledge base, TCM knowledge and experience can be stored during the diagnosis and treatment process, providing fundamental support for TCM data sharing and information exchange, intelligently assisting doctors in diagnosis and treatment, improving the level of TCM diagnosis and treatment, and playing a significant role in the inheritance and development of TCM culture. Currently, such products include Yimaitong Diagnosis and Treatment Knowledge Base, Haodf.com, Yihuwang, and Tencent Medical Dictionary, among other online diagnosis and treatment platforms. However, during the organization of TCM knowledge base resources, numerous problems of non-standard TCM terminology exist, including multiple terms for one meaning, one term with multiple meanings, inconsistent usage habits, obscure disease names, and unclear conceptual relationships between terms. These issues hinder the analysis and mining of TCM diagnosis and treatment knowledge and affect the accuracy and efficiency of machine processing and understanding. Although certain achievements have been made in TCM terminology standardization research, studies specifically oriented toward knowledge base construction in the diagnosis and treatment process are still rare. Based on TCM diagnosis and treatment knowledge bases, this paper establishes the scope of term extraction and selection, constructs the structure of terminology standard files, proposes word meaning control strategies for existing problems, and illustrates the process of knowledge base term extraction and standard file

establishment through specific examples.

2 Literature Review

If the structure and elements of medical cases are compared to the “skeleton” of cases, then TCM terminology constitutes the “flesh and blood.” TCM terminology possesses unique characteristics. For instance, classic TCM works such as *Treatise on Cold Damage* and *Shennong’s Classic of Materia Medica* are heavily colored by classical Chinese. Coupled with China’s vast territory and numerous ethnic groups with diverse dialects, most terms exhibit general and vague features. Zhu Jianping [3] summarized the characteristics of TCM terminology as historical, humanistic, using concrete nouns to express abstract concepts, and qualitative descriptive. As one of China’s national essences, TCM has gradually entered the international arena, and non-standard terminology has become one of the important factors hindering TCM from ascending to the world academic stage [4].

In response to these problems, researchers have conducted extensive work on terminology standardization. First, national, industry, and international standards for TCM terminology have been promulgated. Second, terminology databases have been established with modern information technology. The China Academy of Chinese Medical Sciences has developed a series of TCM databases, totaling 45, containing various standardized data and providing hundreds of thousands of standardized terminology entries, greatly improving the standardization and informatization of terminology [5]. Additionally, to address the low application rate of terminology standards, Guo Yufeng et al. [6] proposed research ideas for constructing a real-world TCM terminology standardization application technology system, pointing out that the main reason for the low application degree of terminology standards in practical scientific research is the lack of standardized application models and tools supporting the standards. To solve the problem of terminology standardization, it is necessary to use information technology and standard regulations to standardize the invocation of existing standardized terminology. Ni Fei et al. [7] used bibliometric methods to conduct standardization research on symptom terminology within the scope of TCM spleen and stomach system diseases. Li Lingqiao and Bao Suzhen [8] clarified the significance of TCM terminology standardization, explained its important role in cultural inheritance, academic exchange, and translation, and proposed that TCM terminology standardization should be achieved through modern thinking and technical methods while retaining its own characteristics and advantages. Zhu Jianping [9] reviewed the achievements and progress of TCM terminology standardization since 2000, pointed out that one of the important factors restricting the modernization of TCM is the standardization of TCM terminology, and put forward corresponding suggestions and prospects for current TCM terminology issues. Xu Ji et al. [10] conducted comparative analysis of several promulgated national TCM terminology standards from aspects of overall situation, word

usage habits, scope of terms, and presentation forms. Wang Chuanchi et al. [11] interpreted relevant policies and news reports on TCM standardization in recent five years from the perspective of current status and existing problems of TCM standardization.

In summary, numerous scholars have mentioned the important role of TCM terminology standardization in TCM development, pointed out existing problems, and proposed corresponding suggestions. In recent years, with the development of information technology, knowledge bases have been widely applied in the medical field. Although experts have published relevant literature on TCM knowledge bases, no scholars have yet conducted research on terminology standardization in TCM diagnosis and treatment knowledge bases. In view of this, this paper takes TCM diagnosis and treatment knowledge bases as the research object, discusses relevant standardization work such as the scope of term selection, standard file structure, and existing problems in the knowledge base, and provides examples as evidence, with the aim of promoting the diversified development of TCM terminology standardization research.

3 Term Extraction and Selection Scope for TCM Diagnosis and Treatment Knowledge Bases

3.1 Based on Existing National and Industry Standards

Since the 1990s, TCM terminology standardization work has received attention and support from national and industry departments. A series of standardization research work on TCM industry terminology has been carried out, and several TCM terminology standards have been promulgated. TCM diagnosis and treatment knowledge bases should first take these as the basic basis. Currently promulgated national standards related to TCM diagnosis and treatment terminology include: *GB/T 20348-2006 Basic Theory Terminology of Traditional Chinese Medicine*, *GB/T 16751.3-1997 Clinical Diagnosis and Treatment Terminology of Traditional Chinese Medicine - Treatment Part*, *GB/T 16751.2-1997 Clinical Diagnosis and Treatment Terminology of Traditional Chinese Medicine - Syndrome Part*, *GB/T 16751.1-1997 Clinical Diagnosis and Treatment Terminology of Traditional Chinese Medicine - Disease Part*, and *GB/T 15657-1995 Classification and Codes of Diseases and Syndromes of Traditional Chinese Medicine*. Industry standards include: *ZY/T 001.1-94 Diagnostic and Therapeutic Criteria for Internal Medicine Diseases of Traditional Chinese Medicine*, *ZY/T 001.2-94 Diagnostic and Therapeutic Criteria for Surgical Diseases of Traditional Chinese Medicine*, *ZY/T 001.3-94 Diagnostic and Therapeutic Criteria for Gynecological Diseases of Traditional Chinese Medicine*, *ZY/T 001.4-94 Diagnostic and Therapeutic Criteria for Pediatric Diseases of Traditional Chinese Medicine*, *ZY/T 001.5-94 Diagnostic and Therapeutic Criteria for Ophthalmological Diseases of Traditional Chinese Medicine*, *ZY/T 001.6-94 Diagnostic and Therapeutic Criteria for Otorhinolaryngological Diseases of Traditional Chinese Medicine*, *ZY/T 001.7-94 Diagnostic and Therapeutic Criteria for Anorectal*

Diseases of Traditional Chinese Medicine, ZY/T 001.8-94 Diagnostic and Therapeutic Criteria for Dermatological Diseases of Traditional Chinese Medicine, and ZY/T 001.9-94 Diagnostic and Therapeutic Criteria for Orthopedic Diseases of Traditional Chinese Medicine, as shown in Table 1 .

3.2 Supplement with Modern Common TCM Terminology

Practice has shown that existing national and industry standards for TCM terminology cannot cover all terms required in TCM diagnosis and treatment research work. In actual TCM diagnosis and treatment operations, the required terms come not only from fixed, standardized terminology but more from practical work, massive clinical diagnosis and treatment, and scientific research. Therefore, in addition to taking national and industry standards for TCM terminology as the basic basis, it is also necessary to supplement with commonly used modern TCM terminology to establish term sets with different functional goals, targeted approaches, and complementary coverage from multiple sources.

For terms not covered by the above standards but needed in actual operations, reference can be made to published and widely recognized TCM dictionaries and textbooks with certain academic standards, such as *Dictionary of TCM Terms*, *TCM Terminology Database*, *Dictionary of Commonly Used TCM Terms*, *TCM Dictionary*, *Great Dictionary of Traditional Chinese Medicine*, *Chinese Traditional Medicine Subject Headings*, *Basic Theory of Traditional Chinese Medicine*, and *Diagnostics of Traditional Chinese Medicine*. Additionally, appropriate TCM terms can be extracted, screened, and organized from well-known, influential TCM journals, famous TCM works, and online medical platforms to provide timely and dynamic updates to the TCM diagnosis and treatment knowledge base term set. To ensure the professionalism and standardization of terms in the TCM diagnosis and treatment knowledge base, these supplementary terms must be consistent with previously entered terms and maintain their standardization. Only after passing the standardization certification mechanism of the TCM diagnosis and treatment terminology set and consulting domain experts can they be officially entered into the TCM diagnosis and treatment knowledge base term set.

4 Structure of TCM Diagnosis and Treatment Knowledge Base Terminology Standard Files

4.1 Basic Structure of TCM Diagnosis and Treatment Knowledge Base

The basic structure of the TCM diagnosis and treatment knowledge base system is mainly divided into three layers: the basic data layer, the rule standard layer, and the diagnosis and treatment plan inference layer, as shown in Figure 1 [Figure 1: see original paper].

The basic data layer includes the medical case database, TCM knowledge base,

and TCM literature database. The medical case database is based on massive TCM diagnosis and treatment cases, with data sourced from electronic medical records of actual clinical diagnosis and treatment patients, obtaining patient consent and removing private information. The medical case words are entered into the knowledge base system after modification according to TCM diagnosis and treatment knowledge base terminology standards. The TCM knowledge base includes detailed disease knowledge, detailed Chinese medicine knowledge, detailed TCM treatment method knowledge, and ancient classic prescriptions. Users can quickly and conveniently learn and query relevant TCM diagnosis and treatment knowledge through subject classification, subject retrieval, or page knowledge indexing [12]. The TCM literature database is based on TCM literature from Wanfang and CNKI, reviewed and tracked by experts. According to the latest research frontiers of each term domestically and internationally, relevant literature is updated in real time, providing the latest and reliable TCM diagnosis and treatment literature resources. Users have the authority to read online and download full texts, providing a basis for clinical diagnosis and treatment decision-making to achieve learning and auxiliary decision-making purposes.

The rule standard layer mainly refers to terminology standards. TCM terminology is based on the latest versions of *Pharmacopoeia of the People's Republic of China* and *General Names of Chinese Medicines* compiled by the National Pharmacopoeia Commission, using generic names rather than commercial names. TCM diagnosis and treatment terminology is based on current national and industry standards, with supplementary terms conforming to the national promulgated *Terminology Work - Principles and Methods (GB/T 10112-1999)* and relevant standards for *Terminology Work - Harmonization of Concepts and Terms*.

The diagnosis and treatment plan inference layer segments the disease information in the diagnosis and treatment knowledge base, establishes an inverted index table, calculates the similarity between queried symptom features and relevant disease content in the diagnosis and treatment knowledge base, and diagnoses disease categories according to similarity ranking. Through user feedback information, diagnosis is further optimized, and Bayesian classification algorithms are used for diagnosis and treatment plan recommendations [13].

4.2 Format Specifications for TCM Diagnosis and Treatment Knowledge Base Terms

The data categories for TCM diagnosis and treatment knowledge base terms must include entry number, term name, and definition, with supplementary information such as corresponding words, notes, and literature references added as needed.

The entry number format adopts a two-layer structure. The first layer is the overall term sequence number, placed at the top column of the entry number,

reflecting the total number of terms in the diagnosis and treatment knowledge base. The second layer is the term code number, composed of a combination of English letters and Arabic numerals, where letters represent the identifier for major term categories and numbers represent the sequence number under the identifier. The code number reflects the hierarchical and relational concepts between terms. For example, the code number for cough disease is BNF010, and its subordinate condition external cough disease is marked as BNF011, and so on.

To prevent confusion between terms, term names adhere to the principle of uniqueness, using only one official name that should be concise and accurate, placed after the entry number. Priority is given to term names from national and industry standards for TCM terminology. In cases of multiple terms for one meaning, alternative terms can be marked as synonyms.

Definitions are placed after the term name on a new line. Definition statements should avoid repetition and contradiction, and should not use demonstrative pronouns such as “this,” “that,” or “the” as the beginning of the definition. Generally, term definitions should include the superordinate concept of the term and explain its distinguishing features from other concepts at the same level. If cited, the source of the definition can be marked in square brackets after the definition [14].

Supplementary information includes corresponding words, notes, and literature references. Corresponding words refer to words with similar or opposite meanings to the term. Notes provide necessary supplementary explanations for term definitions. Literature references list sources related to the term. This supplementary information is placed after the definition and added as needed.

4.3 Categories of TCM Diagnosis and Treatment Knowledge Base Terminology Standard Files

TCM mainly diagnoses diseases through the “four diagnostic methods”: observation, listening/smelling, inquiry, and palpation. Doctors then decide on treatment methods and medications based on specific conditions. The categories of TCM diagnosis and treatment knowledge base terminology standard files should include disease names, symptom manifestations, treatment methods, medication names, and therapeutic criteria.

4.4 Correspondence Between Terminology Standard File Categories and Standards

Terminology standard file categories correspond to national and industry standards for TCM terminology. Disease names correspond to the national standard *GB/T 15657-1995 Classification and Codes of Diseases and Syndromes of Traditional Chinese Medicine*. Symptom manifestations correspond to *GB/T 16751.2-1997 Clinical Diagnosis and Treatment Terminology of Traditional Chinese Medicine - Syndrome Part* and *GB/T 16751.1-1997 Clinical Diagnosis*

and Treatment Terminology of Traditional Chinese Medicine - Disease Part. Treatment methods correspond to *GB/T 16751.3-1997 Clinical Diagnosis and Treatment Terminology of Traditional Chinese Medicine - Treatment Part.* Medication names correspond to *Pharmacopoeia of the People's Republic of China* and *General Names of Chinese Medicines*. Therapeutic criteria correspond to industry standards *ZY/T 001.1-94 Diagnostic and Therapeutic Criteria for Internal Medicine Diseases of Traditional Chinese Medicine*, *ZY/T 001.2-94 Diagnostic and Therapeutic Criteria for Surgical Diseases of Traditional Chinese Medicine*, *ZY/T 001.3-94 Diagnostic and Therapeutic Criteria for Gynecological Diseases of Traditional Chinese Medicine*, *ZY/T 001.4-94 Diagnostic and Therapeutic Criteria for Pediatric Diseases of Traditional Chinese Medicine*, *ZY/T 001.5-94 Diagnostic and Therapeutic Criteria for Ophthalmological Diseases of Traditional Chinese Medicine*, *ZY/T 001.6-94 Diagnostic and Therapeutic Criteria for Otorhinolaryngological Diseases of Traditional Chinese Medicine*, *ZY/T 001.7-94 Diagnostic and Therapeutic Criteria for Anorectal Diseases of Traditional Chinese Medicine*, *ZY/T 001.8-94 Diagnostic and Therapeutic Criteria for Dermatological Diseases of Traditional Chinese Medicine*, and *ZY/T 001.9-94 Diagnostic and Therapeutic Criteria for Orthopedic Diseases of Traditional Chinese Medicine*, as shown in Figure 2 [Figure 2: see original paper].

5 Problems in TCM Diagnosis and Treatment Knowledge Base Terminology Standardization and Word Meaning Control Strategies

5.1 Multiple Terms for One Meaning

TCM terminology is distributed in various standard documents, journals, textbooks, dictionaries, and books, with a huge quantity. Viewed from individual sources, there are no serious problems among terms. However, when terms from multiple sources with different standards are collected, semantic conflicts, intersections, or overlaps may occur, resulting in the phenomenon of multiple terms for one meaning. This manifests specifically in the same concept possibly being represented by different terms in different articles. For example, in TCM ophthalmology diseases, the condition caused by phlegm obstruction in the body, blocked qi and blood, and internal heat, with symptoms of protruding eyeballs and congested, red sclera like a partridge's eye, is termed "partridge eye congealing eye" in the clinical diagnosis and treatment terminology disease part [16], while the diagnostic and therapeutic criteria for TCM ophthalmological diseases use "falcon eye congealing eye" to represent this disease [17]. In fact, "partridge" and "falcon" are two different birds, but "partridge eye congealing eye" and "falcon eye congealing eye" represent the same meaning in the TCM field. The existence of multiple terms for one meaning undoubtedly causes difficulties in selection and great confusion for term users.

A key principle for traditional terminology is that a specific concept should be represented by only one term, i.e., terms should have uniqueness. In response to

this situation, the strategy is to extract all words representing the same meaning from different sources, then use word segmentation software for word frequency statistics. If the word frequencies differ significantly, the term with the highest frequency across all sources should be selected as the only standardized official term for that concept in the TCM diagnosis and treatment knowledge base [18-19]. Furthermore, the goal of terminology standardization means that the purpose of selecting terms and entering them into the database is to recommend the most acceptable term for a specific concept. Other terms representing the same concept should be listed separately, with each synonym assigned a certain standardized status: admitted term, non-recommended term, or obsolete term.

5.2 One Term with Multiple Meanings

TCM has a long history, forming a medical theoretical system through continuous summarization from thousands of years of practical experience of ancient people fighting diseases. Additionally, China has a vast territory with 56 ethnic groups. Due to the limitations of dialects and regional isolation, information exchange was inconvenient, inevitably leading to the problem of one term with multiple meanings in TCM terminology. For example, the term “life gate” has three meanings: it represents an acupoint name (the life gate point of the governor vessel), located opposite the navel at the lumbar spine; it can also represent a human tissue organ, referring to the kidneys in the five-zang theory and to the eyes in *Huangdi Neijing*; it also symbolizes the root of human life and vital energy [20]. Additionally, the TCM term “qi” can be understood in six different ways: “energy,” “air,” “substance,” “activity,” “symptom manifestation,” and “mental state.”

TCM scholar Xu Zhiqian once statistically analyzed medical terms in *Great Dictionary of Traditional Chinese Medicine - Basic Theory Volume* and found that the ratio of one term with multiple meanings in this book reached 15.8%, greatly exceeding the 0.3% ratio in Western medicine [21]. One term with multiple meanings violates the “law of identity” principle of standardized terminology, causing unclear concepts in TCM terminology. This 无形中 increases the difficulty for TCM learners to distinguish word meanings, seriously restricts TCM academic exchange, hinders the inheritance and development of TCM, and constitutes a major stumbling block to TCM terminology standardization.

For the phenomenon of one term with multiple meanings, all definitions of the term can first be listed, uncommon definitions can be deleted, and then different meanings can be differentiated. By adding annotations after the term name, the term can correspond to different meanings. For difficult cases of one term with multiple meanings that are hard to differentiate, the Delphi method can be adopted. This is an effective expert consultation method widely used in the TCM field. First, an expert group of TCM scholars and famous doctors with certain professional standards is formed. Then, consultation forms are compiled based on the problem, and multiple rounds of anonymous consultation surveys are conducted. The collected opinions are continuously organized, summarized,

and statistically analyzed to finally obtain results, adopting definitions that are widely recognized, easy to understand, and scientifically reasonable [22]. For example, the term “throat fossa” originally had two meanings: one refers to the suprasternal fossa on the human body, and the other refers to the depression between the pharyngeal arch and palatoglossal arch. After research and discussion by experts, the second meaning was adopted as the TCM definition of “throat fossa,” discarding the meaning of suprasternal fossa [23].

5.3 Inconsistent Usage Habits

Inconsistent usage habits exist among multiple terminology standards, including different suffix usage habits for disease names, different term selection, and different character usage. These problems prevent the unification of TCM terminology usage and constrain the process of terminology standardization.

(1) Disease Name Suffixes. Different terminology standards often have inconsistent usage habits regarding whether to add the suffix “disease” or “syndrome” to disease names. Taking the national standards *GB/T 15657-1995 Classification and Codes of Diseases and Syndromes of Traditional Chinese Medicine* and *GB/T 16751.1-1997 Clinical Diagnosis and Treatment Terminology of Traditional Chinese Medicine - Disease Part* as examples, the former always adds the suffix “disease” after disease names, while the latter sometimes does not add the suffix “disease,” naming diseases directly. For instance, lung failure disease, lung atrophy disease, and chest bi-heart pain disease in the former are expressed as lung failure, lung atrophy, and chest bi-heart pain in the latter standard.

(2) Character Usage. When prescribing, TCM doctors, due to patients waiting, have developed the habit of writing quickly and carelessly, often using simplified characters or homophones according to personal habits. Over time, this has caused different character usage habits among TCM terminology standards. For example, “milk fistula,” “drum distension,” “scar,” and “extinguishing wind” in *Clinical Diagnosis and Treatment Terminology of Traditional Chinese Medicine* correspond to “milk leakage,” “tympanites,” “scar,” and “calming wind” in *TCM Terminology* and *Classification and Codes of Diseases and Syndromes of Traditional Chinese Medicine* [10].

The problem of inconsistent usage habits in TCM terminology urgently requires further standardization. First, it is necessary to strengthen the professional ethics and academic literacy education of doctors and TCM scholars, making them fully aware of the confusion and losses caused by using non-standard terminology, correcting the habit of arbitrarily using terminology, familiarizing themselves with standardized TCM terminology expression, and developing good habits. Additionally, for similar inconsistencies in terminology usage among national and industry standards, it is recommended that relevant TCM departments regularly revise promulgated standards, conduct unified standardization processing of terminology, and adopt only one term form to represent, avoiding

the phenomenon of two different representation methods for the same term. The usage habits of terminology in TCM diagnosis and treatment knowledge bases should first refer to national and industry standards. If conflicts exist, *Chinese Traditional Medicine Subject Headings* should be used as the basis. For non-standardized TCM terms not yet included in existing national and industry standards and *Chinese Traditional Medicine Subject Headings*, consistency in the TCM diagnosis and treatment knowledge base should be ensured, strictly implementing one standard and avoiding multiple expressions for one term.

5.4 Obscure and Difficult-to-Understand Disease Names

Inherited from ancient times, TCM texts bear characteristics of classical Chinese. Some TCM terms contain relatively obscure characters, with unknown pronunciations, such as “scrofula,” “masses,” “convulsions,” “羸,” and “燠.” Some TCM terms have simple individual characters, but the combined meaning has little apparent connection with the disease and is difficult to understand. For example: “crouching rabbit abscess” refers to toxic sores on the thigh caused by stomach fire; “correct water” refers to edema, hypertension, and oliguria caused by wind-heat dampness toxicity; “fat qi” refers to accumulation-type diseases with abdominal swelling and bleeding due to internal blood stasis; “crouching beam” refers to diseases with qi and blood obstruction, abdominal discomfort, and right lower abdominal mass caused by intestinal filth toxicity.

Obscure and difficult-to-understand disease names affect people’s understanding and exchange of TCM academic knowledge and hinder TCM from entering the international community. If these TCM terms cannot be understood by many medical students or even diagnosis and treatment doctors, how can they explain them to patients? How can patients and society understand them? How can TCM go global? These obscure TCM terms not only fail to reflect the etiology, symptoms, and characteristics of diseases but also create obstacles to understanding, let alone guiding and analyzing with these terms.

Considering that TCM terminology differs from general language, involving both TCM professional knowledge systems and language systems, when dealing with obscure and difficult-to-understand disease name terminology, comprehensive consideration of both linguistic and professional factors should be taken into account [24]. TCM terminology standardization work requires cooperation and exchange between TCM scholars and linguists. Based on a deep understanding of the internal meaning and characteristics of terms, obscure TCM terms should be given new words appropriate to the times to ensure accuracy and simplicity. TCM cannot remain unchanged or view itself in an isolated and static manner. It is necessary to transform obscure TCM terms, develop TCM through inheritance, make TCM continuously modernized, understood and accepted by more people, and expand its influence.

5.5 Unclear Concepts and Relationships Between Terms

Limited by objective conditions in ancient times, TCM diagnosis methods were primarily observation, listening/smelling, inquiry, and palpation. However, when facing the difficulty of understanding relationships between diseases, problems of unclear term concepts and hierarchical confusion arise. For example, people's understanding of the terms "syndrome," "symptom," "sign," and "syndrome manifestation" has always had differences, with frequent mixed usage of these terms. Additionally, terms listed earlier are sometimes parallelly included under their lower-level terms. These different understandings of term concepts make it difficult for TCM scholars to reach consensus, directly affecting subsequent standardization work such as specific disease naming and classification, as well as academic exchange.

Concepts should have relationships between concepts and characteristics, i.e., attributes in the ontological knowledge system [25]. The connotation of a term reflects the sum of its essential attributes, while the extension is the category reflected by the term concept, i.e., the scope to which the term concept applies. Therefore, defining a TCM term concept requires clarifying its connotation and extension, constructing a clear conceptual system, and thereby distinguishing these terms from the source. Constructing a coherent TCM term conceptual system requires arranging all TCM terms in the TCM diagnosis and treatment database according to clear superior-subordinate relationships, indicating the logical relationships between terms so that they form a clear hierarchical system. Taking the TCM disease term "cough disease" as an example, its connotation is a lung system disease caused by external pathogenic factors and other causes, with lung failing to purify and descend, qi rushing through the respiratory tract, and producing cough sounds. Its hyponyms are external cough disease and internal cough disease, with external cough disease caused by external pathogenic factors and internal cough disease caused by lung organ dysfunction. The coordinate terms of cough disease include lung failure disease, lung leakage disease, and asthma disease, whose hypernym is lung system disease category, all belonging to diseases appearing in the lung, as shown in Figure 3 [Figure 3: see original paper].

6 Methods and Case Analysis of Term Extraction and Standard File Establishment for TCM Diagnosis and Treatment Knowledge Bases

6.1 Multiple Terms for One Meaning

6.1.1 Method and Process TCM knowledge itself has certain fuzziness and tacit understanding. Influenced by historical factors, ancient people's understanding of TCM concepts was not specific and could only be experienced and imagined through actual cases around them, which also caused possible differences in each person's understanding of TCM, resulting in the phenomenon of multiple terms for one meaning. Multiple terms for one meaning, i.e., syn-

onym problems, are common in TCM terminology, causing confusion in term usage and understanding, which is not conducive to terminology standardization. Building TCM diagnosis and treatment terminology standard files must solve this problem, adhering to the law of identity principle in terminology standardization by selecting only one term as the official standardized term in the TCM diagnosis and treatment knowledge base. The process for establishing standard files under multiple terms for one meaning is shown in Figure 4 [Figure 4: see original paper].

First, according to the above scope of term extraction and selection for TCM diagnosis and treatment knowledge bases, standard file establishment should first take existing national and industry standards for TCM terminology as the basic basis. In cases of multiple terms for one meaning, if national and industry standards for TCM terminology have a unified judgment on the standard term among multiple terms, that term can be directly determined as the standardized term for the TCM diagnosis and treatment knowledge base, with other terms marked as non-recommended or obsolete terms. Second, since the operation for multiple terms for one meaning is relatively difficult compared to other problems, an important step in handling this issue is to determine whether multiple terms express the same meaning, which is the foundation for further progress. Therefore, for TCM terms that cannot be based on national and industry standards (such as when multiple standards have disputes about whether multiple terms are synonyms, or when commonly used modern TCM terms not yet included in national and industry standards), it is necessary for ten experts in the TCM field to analyze multiple terms. If most experts believe they constitute synonyms, word frequency statistics after word segmentation can be used, with the term used more frequently in core TCM journals selected as the standardized term for the TCM diagnosis and treatment knowledge base. If opinions are divided and most experts believe there are significant differences between multiple terms that cannot constitute synonyms, then multiple terms are not treated as synonyms but listed as two independent terms. Here, this paper takes the TCM terms “nutritive qi” and “glory qi,” as well as “dampness-heat accumulating in spleen” and “spleen channel dampness-heat,” as specific examples to illustrate term extraction and standard file establishment under multiple terms for one meaning.

6.1.2 Term Extraction and Standard File Establishment—Case of “Nutritive Qi” and “Glory Qi” Baidu Baike and Medical Encyclopedia annotate the term “glory qi” as: glory qi is nutritive qi, meaning that glory qi means nutritive qi. Both “nutritive qi” and “glory qi” are regarded as synonyms, referring to qi circulating in the vessels with nutritional functions. “Nutritive qi” and “glory qi” often appear in many TCM journals and papers, with mixed usage still occurring today. Thematic retrieval of the term “nutritive qi” on CNKI yields 350 results, with the latest TCM literature containing this term published in July 2020. Retrieval of the term “glory qi” yields 25 results, with the latest TCM literature containing this term published in April 2019.

According to the above term extraction principles for TCM diagnosis and treatment knowledge bases, existing national and industry standards for TCM terminology should first be used as the basic basis. If not available, the formal term for the TCM diagnosis and treatment knowledge base should be determined based on their frequency of occurrence in core TCM journals. After retrieval, “nutritive qi” appears 16 times in national TCM terminology standards, while “glory qi” appears 0 times. Therefore, “nutritive qi” is selected as the formal term for the TCM diagnosis and treatment knowledge base, with “glory qi” marked as a non-recommended or obsolete term.

6.1.3 Term Extraction and Standard File Establishment—Case of “Dampness-Heat Accumulating in Spleen” and “Spleen Channel Dampness-Heat” In *GB/T 16751.2-1997 Clinical Diagnosis and Treatment Terminology of Traditional Chinese Medicine - Syndrome Part*, “dampness-heat accumulating in spleen” and “spleen channel dampness-heat” are synonyms, both referring to dampness-heat syndrome caused by dampness-heat obstruction, spleen failing to transform and transport, resulting in abdominal distension, nausea, poor appetite, heavy limbs, unsatisfactory loose stools, or yellowing of face and eyes, body heat not spreading, sweating without heat resolution, thirst without excessive drinking, red tongue with yellow greasy coating, and soft pulse [26], where “dampness-heat accumulating in spleen” is the subject term and “spleen channel dampness-heat” is listed as its synonym. However, in *GB/T 15657-1995 Classification and Codes of Diseases and Syndromes of Traditional Chinese Medicine*, “spleen channel dampness-heat” is the subject term, and the term “dampness-heat accumulating in spleen” does not appear. Comparing the two, the two national TCM terminology standards have inconsistent standards. In this case, national TCM terminology standards obviously cannot be used as the basis for determining which of “dampness-heat accumulating in spleen” and “spleen channel dampness-heat” should be the standardized term for the TCM diagnosis and treatment knowledge base. Therefore, the next step is for ten experts in the TCM field to analyze “dampness-heat accumulating in spleen” and “spleen channel dampness-heat” to determine whether they are synonyms. The author consulted more than ten clinical TCM experts from the Traditional Chinese Medicine Department of Beijing Tongren Hospital, Beijing University of Chinese Medicine, Anorectal Department of Shenzhen Traditional Chinese Medicine Hospital, Gastroenterology Department of the First Clinical Hospital of Jilin Academy of Traditional Chinese Medicine, Urology Department of Changchun Traditional Chinese Medicine Hospital, Lung Disease Department of Jilin Academy of Traditional Chinese Medicine, Jiangxi Provincial Hospital of Traditional Chinese Medicine, Traditional Chinese Medicine Department of Jiangxi Hukou Maternal and Child Health Hospital, Affiliated Hospital of Liaoning University of Traditional Chinese Medicine, and Traditional Chinese Medicine Department of Zibo Municipal Hospital in Shandong Province. Due to significant controversy among experts about whether “dampness-heat accumulating in spleen” and

“spleen channel dampness-heat” are synonyms, they cannot be determined as synonyms. Therefore, in the TCM diagnosis and treatment knowledge base, “dampness-heat accumulating in spleen” and “spleen channel dampness-heat” are not treated as synonyms but listed as two independent terms.

6.2 One Term with Multiple Meanings

6.2.1 Method and Process Chinese culture is extensive and profound, with rich meanings in characters. The situation of one term with multiple meanings often occurs, and TCM, with its long development history, is no exception. One term with multiple meanings in TCM terminology may mean that the same term has different meanings in different fields. For example, the term “numbness” in the literary field refers to spiritual indifference, slow thinking, and dull response, while in the medical field it refers to physical loss of sensation. It may also mean that one term has multiple different interpretations in the same field. For example, “triple burner” has three meanings in TCM: one refers to one of the six fu organs, collectively called upper, middle, and lower burners; two refers to ear acupoint names; three refers to massage area names. One term with multiple meanings makes it difficult for TCM diagnosis and treatment knowledge bases to select definitions, easily causing semantic confusion and understanding difficulties, which is not conducive to terminology standardization. Therefore, for the phenomenon of one term with multiple meanings, the method and process for establishing standard files for TCM diagnosis and treatment knowledge bases is proposed, as shown in Figure 5 [Figure 5: see original paper].

For the problem of one term with multiple meanings, first list all definitions of the term and divide them by field. Then determine whether these definitions belong to the TCM field. Since the purpose is to establish TCM diagnosis and treatment knowledge base terminology standard files, definitions outside the TCM field can be deleted, retaining only the required TCM definitions. If only one TCM definition remains after deletion, that term definition can be directly determined. If multiple TCM definitions still exist, meanings can be differentiated by adding annotations after the term name, enabling one meaning to correspond to one term, thereby achieving one term with one meaning. Below, the TCM term “triple burner” is used as an example.

6.2.2 Term Extraction and Standard File Establishment—Case of “Triple Burner” “Triple burner” has three different definitions. The first definition refers to one of the six fu organs of the human body, collectively called upper, middle, and lower burners. The second definition refers to an ear acupoint name, the triple burner ear acupoint, located behind and below the external ear door, between the lung and endocrine points. The third definition is the name of a massage area, with four positions used for strengthening the spleen and stomach and treating digestion problems. Since these three definitions all belong to the TCM field, it is necessary to differentiate different meanings. When entering the term name, marking “triple burner” with “ear

acupoint name” after the term, such as “triple burner (ear acupoint name),” can correspond to the second definition. Similarly, marking “triple burner” with “massage area name” after the term, such as “triple burner (massage area name),” corresponds to the third definition. “Triple burner” without annotation corresponds to the most widely used first definition. Through this method, one term with multiple meanings can be transformed into one term with one meaning.

6.3 Unclear Term Concepts and Relationships

6.3.1 Method and Schematic Diagram Term concepts reflect the definitions and connotations of terms, while term relationships reflect the hierarchy and logic between terms. Both are important components of building TCM diagnosis and treatment knowledge base terminology sets. Unclear term concept expression and chaotic relationship representation can easily cause information loss or redundancy, leading to users’ inability to accurately obtain effective information and causing understanding difficulties. Therefore, this problem should be given due attention. The method for establishing TCM diagnosis and treatment knowledge base term concepts and relationships is introduced below.

(1) Concept Composition. When setting TCM diagnosis and treatment knowledge base term concepts, TCM terminology characteristics should be followed, consisting of the term’s relational features and attribute features. Relational features here refer to the term’s hypernym, indicating which category and scope the term belongs to. Attribute features describe the relevant attributes and characteristics of the term. If relational features are vertical concept connections, attribute features are horizontal concept connections, jointly defining the term concept.

(2) Term Relationship Establishment. Terms in TCM diagnosis and treatment knowledge bases are mainly divided into five categories: disease names, symptom manifestations, treatment methods, therapeutic criteria, and medication names, with reference standards detailed in section 3.4 above. Therefore, when constructing term relationships, the approach is to radiate from disease names as the origin, supplementing with hyponyms, hypernyms, and relative terms of diseases based on the above five term categories. Relative terms refer to synonyms, similar terms, and antonyms, as shown in Figure 6 [Figure 6: see original paper].

6.3.2 Term Extraction and Standard File Establishment—Case of “Wheezing Disease” **(1) Concept.** The national standard for TCM diseases defines “wheezing disease” as: a paroxysmal lung system disease usually caused by external pathogenic factors, emotional injury, and irregular diet, clinically characterized by sudden respiratory difficulty and wheezing sounds in the throat [27]. Therefore, the conceptual definition of “wheezing disease” can be recorded as: (hypernym) lung system disease category; (etiology) external

pathogenic factors, emotion, diet; (clinical manifestations) sudden respiratory difficulty, wheezing sounds in the throat.

(2) Term Relationship Establishment. Based on the above schematic diagram, the relationship diagram for wheezing disease is established, as shown in Figure 7 [Figure 7: see original paper]. Here, the disease name, hypernym, and symptom manifestations of “wheezing disease” refer to the national standard - classification and codes of TCM diseases and TCM clinical diagnosis and treatment terminology disease part. Therapeutic criteria refer to the industry standard - diagnostic and therapeutic criteria for internal medicine diseases of traditional Chinese medicine. Treatment methods refer to the national standard - TCM clinical diagnosis and treatment terminology treatment part. Medication names refer to Chapter 2, Section 2 on wheezing disease in *Internal Medicine of Traditional Chinese Medicine*. The medication names here are only general drugs used for treating “wheezing disease,” not prescription recommendations. Specific medications to be taken should be consulted with physicians based on individual conditions.

References

- [1] State Council. State Council Issues “Strategic Planning Outline for the Development of Traditional Chinese Medicine (2016-2030)” [N]. People’s Daily, 2016-02-27(1).
- [2] Qiu Junping, Chen Jingquan. Knowledge Warehouse and Its Application in Enterprise Management [J]. Information Studies: Theory & Application, 2003(4): 324-326, 323.
- [3] Zhu Jianping. Standardization of TCM Terminology and Modernization and Internationalization of TCM [J]. China Journal of Traditional Chinese Medicine and Pharmacy, 2006(1): 6-8.
- [4] Dong Yan, Li Haiyan, Cui Meng, et al. Overview and Improvement Measures for the Construction of TCM Clinical Terminology System [J]. Journal of Medical Informatics, 2014, 35(8): 43-48.
- [5] Cui Meng, Li Haiyan, Yang Shuo, et al. Research Progress in the Theoretical Scientific Field of TCM Informatics [J]. Chinese Journal of Library and Information Science for Traditional Chinese Medicine, 2015, 39(5): 1-6.
- [6] Guo Yufeng, Xie Qi, Zhou Xiaji, et al. Thoughts on Constructing a Real-World Standardized Application Technology System for TCM Clinical Terminology [J]. Journal of Traditional Chinese Medicine, 2015, 56(7): 557-560.
- [7] Ni Fei, Qu Jinqiao, Wang Caixia, et al. Standardization Research on Symptom Terminology of TCM Spleen and Stomach Diseases Based on Bibliometric Methods [J]. Liaoning Journal of Traditional Chinese Medicine, 2018, 45(10): 2086-2087.

- [8] Li Lingqiao, Bao Suzhen. On the Standardization of TCM Terminology [J]. Journal of Shaanxi University of Chinese Medicine, 2015, 38(4): 14-16.
- [9] Zhu Jianping. Practice and Reflection on the Standardization of TCM Terminology [J]. China Terminology, 2017, 19(6): 11-14.
- [10] Xu Ji, Shi Yi, Yuan Min, et al. Comparative Study on National Standards of TCM Terminology [J]. Lishizhen Medicine and Materia Medica Research, 2015, 26(9): 2294-2295.
- [11] Wang Chuanchi, Yang Yan, Hu Jingqing, et al. Current Status, Problems and Related Policy Interpretation of TCM Standardization Research and Analysis [J]. China Journal of Traditional Chinese Medicine and Pharmacy, 2018, 33(9): 3776-3780.
- [12] Zhang Yueyue, Xu Xin. Design and Construction of Dietary Therapy Knowledge Base for Sub-Healthy Population Based on Subject Headings [J]. Library and Information Service, 2015, 59(14): 6-14.
- [13] Ma Yu, Wang Hongzhi, Gao Hong. Auxiliary Diagnosis and Treatment System Based on Medical Knowledge Base [J]. Journal of Computer Applications, 2016, 36(S1): 217-219, 261.
- [14] Hong Wei. Review and Editing Points for Terminology and Terminology Standards [J]. China Quality and Standards Review, 2017(8): 44-46.
- [15] Fu Zhu, Wang Yuefen. Research on Several Issues in the Term Collection Stage of Co-word Analysis [J]. Journal of the China Society for Scientific and Technical Information, 2016, 35(7): 704-713.
- [16] State Bureau of Quality and Technical Supervision. GB/T 16751.1-1997 Clinical Diagnosis and Treatment Terminology of Traditional Chinese Medicine - Disease Part [S]. Beijing: Standards Press of China, 1997.
- [17] State Administration of Traditional Chinese Medicine. ZY/T 001.5-1994 Diagnostic and Therapeutic Criteria for Ophthalmological Diseases of Traditional Chinese Medicine [S]. Beijing: Standards Press of China, 1995.
- [18] Wang Liuan, Chang Chun. Research on Processing Methods of Abbreviations in Synonymous Term Merging [J]. Library and Information Service, 2014, 58(2): 121-125.
- [19] Chang Chun. Determining Concept Attributes of Thesauri Based on Word Frequency Information [J]. Library and Information Service, 2013, 57(16): 11-14, 24.
- [20] Su Li. One Term with Multiple Meanings (I) [J]. Shaanxi Journal of Traditional Chinese Medicine, 1981(2): 48.
- [21] Xu Zhiquan. Polysemy and Standardization of TCM Terminology [J]. Journal of Shandong University of Traditional Chinese Medicine, 1994, 18(5): 329.

- [22] Dong Xinglu, Cao Kegang, Ma Bin, et al. Analysis of Real-World TCM Terminology Research in the Big Data Era [J]. China Journal of Traditional Chinese Medicine and Pharmacy, 2014, 29(9): 2724-2726.
- [23] TCM Terminology Committee. TCM Terminology (2013) [M]. Beijing: Science Press, 2014.
- [24] Liu Wanghua, Li Hua, Zhou Xiaoqing. Thoughts on the Standardization of TCM Disease Names—Learning from Clinical Diagnosis and Treatment Terminology of Traditional Chinese Medicine [J]. Guangming Journal of Chinese Medicine, 2002(1): 5-7.
- [25] Chang Chun. Discussion on Concept Term Selection and Inter-term Relationship Establishment from “Cow Leg” [J]. Journal of Intelligence, 2009, 28(10): 24-27.
- [26] State Bureau of Quality and Technical Supervision. GB/T 16751.2-1997 Clinical Diagnosis and Treatment Terminology of Traditional Chinese Medicine - Syndrome Part [S]. Beijing: Standards Press of China, 1997.
- [27] State Bureau of Quality and Technical Supervision. GB/T 16751.1-1997 Clinical Diagnosis and Treatment Terminology of Traditional Chinese Medicine - Disease Part [S]. Beijing: Standards Press of China, 1997.
- [28] Zhang Pan, Mao Shusong, Deng Wenping. Research on Standardization of TCM Information Standard Terminology [J]. Journal of Medical Informatics, 2018, 39(5): 12-15.

Author Contributions

Ma Jie: Proposed the research proposition and overall framework, revised and finalized the manuscript.

Wu Wenling: Responsible for manuscript writing and revision.

Sun Hengyu: Responsible for manuscript writing and revision.

Wang Chunxiu: Responsible for manuscript writing and revision.

Ma Jie^{1,2}, Wu Wenling¹, Sun Hengyu³, Wang Chunxiu¹

¹School of Management, Jilin University, Changchun 130022

²Information Resources Research Center, Jilin University, Changchun 130022

³The First Clinical Hospital of Jilin Academy of Traditional Chinese Medicine, Changchun 130022

Study on the Standardization of Terms in TCM Diagnosis and Treatment Knowledge Base

Abstract: [Purpose/Significance] TCM terminology serves as the carrier of TCM knowledge and constitutes a vital component of TCM diagnosis and treatment knowledge bases. Research on terminology standardization for TCM di-

agnosis and treatment knowledge bases helps improve the quality of medical case data information in these knowledge bases and promotes TCM culture. [Method/Process] This study determines the scope of term extraction and selection for TCM diagnosis and treatment knowledge bases, constructs the structure of terminology standard files, proposes word meaning control strategies for existing problems in terminology standardization, and illustrates the process of knowledge base term extraction and standard file establishment through specific examples. [Result/Conclusion] Terminology standardization based on TCM diagnosis and treatment knowledge bases can enhance the utilization efficiency of medical cases in these knowledge bases and advance the modernization and internationalization of TCM.

Keywords: TCM diagnosis and treatment knowledge base; TCM terminology; standardization of terms

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv — Machine translation. Verify with original.