

A Framework Design of Mongolian Idioms Knowledge Base

Hai Yinhua

Abstract—Building a "Knowledge Base of Mongolian Idioms" is not only the effective way of protecting, developing and utilizing Mongolian language resources, but also providing formal knowledge for machine translation, corpus processing, text proofreading, and other fields which can meet the urgent needs of Mongolian Information Processing. Developing a teaching-aid program will provide a convenient tool for the Mongolian students of different ages and at different levels. The aim of this paper is using the computational linguistic theory and method to developing a "Knowledge Base of Mongolian Idioms". This paper will dissertate a framework of the knowledge base from the perspective of the scale and structure, attribute field, management software, and the analysis of application prospect and etc.

Index Terms—Mongolian, knowledge base of idioms, framework design.

I. THE URGENCY OF DEVELOPING KNOWLEDGE BASE OF MONGOLIAN IDIOMS

In recent years, the research of idioms, fixed phrases for natural language understanding has become a hot topic in information processing field. This research is a part of the project of Inner Mongolia University about building language knowledge base supported by National Social Science Fund from 2012. Since 1980s, the Mongolian information processing has made certain accomplishments in basic research and application. As a special element of multi-level and complicated Mongolian language system, Mongolian idioms, ranging from word processing to sentence processing, have been ignored. Much work in the digitization of Mongolian idioms needs to be done.

As an essential part of the Mongolian language resources, idioms have run a long course and carried a lot of cultural heritages. It can reflect the delicate feelings, interests, values and thinking styles of Mongolian people. Also, it provides valuable resources for the study of different aspects of Mongolian civilization. Hence, in order to get a better protection and utilization, detailed study and further research in the exploitation, development and collection of the Mongolian idioms are needed. The rapid development of information age has brought a great challenge to the vocabulary and application of Mongolian idioms. Mongolian

linguist, Professor Nasun-urt held the knowledge base of idioms as an important language resources construction in his "the Presumption of Mongolian Language Resource Platform Framework" [1]. In addition, we discussed in detail the importance and urgency of constructing and using Mongolian idioms knowledge base in another text [2].

Human-oriented, traditional study of Mongolian idioms with multiple methods, from the multi perspective carried on the multi-dimensional research. The results of those studies provide a reference for our topic. Wood-block version of Mongolian idioms lexicography was developed in Qing Dynasty in the late 19th century, and the version lasted for more than 200 years. Nowadays, some reference dictionaries, such as *Brief Dictionary of Mongolian Idioms* (2000) [3], *Dictionary of Mongolian Idioms* (2001) [4], have met the various needs of people to a large extent. However, from the perspective of information processing, with the change of serving target, certain contents of the printed dictionaries cannot be directly applied to the machine dictionary. Particularly, the information like classification and interpretation of individual understanding are not applicable to the machine dictionary.

Previously, scholars have in-depth studies on formalization issues related to grammatical features of Mongolian fixed phrases for computer understanding. The representative works are *Dictionary of Modern Mongolian Fixed Phrases with Detailed Description of Grammatical Information* (2005)[5] and *A study on compound words for Mongolian information processing*[6]. However, Mongolian idioms cover a wide range, either in forms of phrases and sentences, or in different types, and with quite a number of sub-categories; including not only *xoršiy_a üge*¹ (ᠬᠣᠷᠰᠢᠶ᠋ᠠ ᠤᠭᠡ), *xebšil üge* (ᠬᠡᠪᠰᠢᠯ ᠤᠭᠡ), *xabutu üge* (ᠬᠠᠪᠲᠤ ᠤᠭᠡ), *togtamal üge* (ᠲᠣᠭᠲᠠᠮᠠᠯ ᠤᠭᠡ), *ner_atomiy_a* (ᠨᠡᠷ ᠠᠲᠣᠮᠢᠶ᠋ᠠ), but also *jür üge* (ᠵᠦᠷ ᠤᠭᠡ), *sečen üge* (ᠰᠡᠴᠡᠨ ᠤᠭᠡ), *xošong üge* (ᠬᠣᠰᠣᠩ ᠤᠭᠡ), *ceger üge* (ᠴᠡᠭᠡᠷ ᠤᠭᠡ), *surgical üge* (ᠰᠤᠷᠭᠠᠯ ᠤᠭᠡ), *lorjoo üge* (ᠯᠣᠷᠵᠠ ᠤᠭᠡ) and other sub-categories.

Essentially these sub classification and fixed phrases belong to different level. At present, a lot of researches related on Mongolian idioms still lingering on a blank stage, such as classification, clarification of sub-categories, and semantic features of sentence-formed idioms, especially the research on the semantic properties of the machine-understanding-oriented whole idioms. The relevant research of Mongolian idioms for Natural Language Understanding is underdeveloped, which not only directly affects the corpus processing, text classification, machine translation, information retrieval and a lot of other current

Manuscript received April 18, 2014; revised May 29, 2014. This research is based on the project of NSSF grant of China "the development and research of the knowledge base of Mongolian idioms"(No.12CYY062), the project of Key NSF grant of China-"the theories and methods research of multi-ethnic character recognition and understanding" (No.61032008).

Hai Yinhua is with the School of Mongolian Studies, Inner Mongolia University, Hohhot,010021,China (e-mail: haiyh2008@163.com).

¹ Mongolian notes in this paper is in the form of the International Phonetic Alphabet.

undergoing research work, but also may cause a lot of barriers to the further research and application in the future.

II. THE SCALE AND STRUCTURE OF KNOWLEDGE BASE

Conceiving the overall framework of Mongolian idioms knowledge base is based on the theory of lexicology and lexicography, will collect common idioms as the source of entries from *One Million Word Mongolian Corpus*² and some current people-oriented idiom dictionaries, such as *The Dictionary of Mongolian Idiom*³, *Dictionary of Mongolian language*(1997) [7], *Mongolian Dictionary with Detailed Explanation* (2008) [8] to establish the knowledge general base of Mongolian idioms(the sample of knowledge base of Mongolian idioms as shown in Fig. 1).

Up to now, the knowledge base has collected 22000 idioms in total. To clear up ambiguity, several branch base and supplementary base will be established to represent idioms of synonymy or antonym, with single or polysemy, same or similar, reverse or opposite meanings, providing multiple solutions for auto-processing semantic analysis.

A. The Explanation of Attributes Field of General Base

- 1) Number (No): Automatically generated digital numbers involve the number of specific idiom entries in the knowledge base resources.
- 2) Entry (HELECE): fill in the Latin transformed script of Mongolian idioms. For example, fill in "NAMVR-VN EBESU HABVR-VN ALTA" in the corresponding field of Mongolian script "ᠨᠠᠮᠪᠦᠷᠦᠨᠢᠪᠢᠰᠤᠬᠠᠪᠪᠦᠷᠦᠨᠠᠯᠲᠤ" instead.

First, as a blueprint, our main resource--- *the Dictionary of Mongolian Idioms* (2001), we artificially selected the common idioms from it. Secondly, relying mainly on the program and supported by our artificial selection, we include all idioms from *One Million Word Mongolian Corpus* with POS Tagging, and make sure to exclude double counting.

- 3) Sub-category (BTOROL): fill the information marks of grammatical classification as the sub-category codes, for example, fill in "R" in the corresponding field of "ᠨᠠᠮᠪᠦᠷᠦᠨᠢᠪᠢᠰᠤᠬᠠᠪᠪᠦᠷᠦᠨᠠᠯᠲᠤ", indicating "jür üge" (proverb).

HELECE	MONGOLIAN	BTOROL	HYBL	HYB2	HYB3
A USUG ERDEM-VN DEGEJI AYAG A CAI IDEGEN-U DEGEJ	ᠠᠤᠰᠤᠭᠡᠷᠢᠳᠡᠮᠦᠨᠳᠡᠭᠡᠵᠢᠠᠶᠠᠭᠠᠴᠠᠢᠢᠳᠡᠭᠡᠨᠤᠳᠡᠭᠡᠵᠢ				
A USUG-YN TOLOGAI MEDEHI UGEI	ᠠᠤᠰᠤᠭᠡᠨᠲᠣᠯᠣᠭᠠᠢᠮᠡᠳᠡᠬᠢᠤᠭᠡᠢ	H			USUG BICIᠭᠡᠯᠢ
A USUG NOM-VN DEGEDU ABV EJI HOMON-U DEGEDU	ᠠᠤᠰᠤᠭᠡᠨᠠᠮᠤᠪᠦᠷᠦᠨᠳᠡᠭᠡᠳᠤᠠᠪᠪᠦᠷᠦᠨᠠᠯᠲᠤᠳᠡᠭᠡᠳᠤ	G			USUG BOL ERD
A MA UGEI	ᠠᠮᠠᠤᠭᠡᠢ	Y			USUG BICIᠭᠡᠯᠢ
A LAMA	ᠠᠯᠠᠮᠠ	Y			ᠵᠢᠯᠢᠲᠣᠭᠣᠰᠠᠷ
A-YI JIGAGAD E-YI BARIYARV	ᠠᠶᠢᠵᠢᠭᠠᠭᠠᠳᠡᠶᠢᠪᠠᠷᠢᠶᠠᠷᠦᠪ	H			BITEGUI HECEᠭᠡ
A-VIN ACVG UGEI ARDA-VIN TEG UGEI	ᠠᠶᠢᠨᠠᠴᠢᠪᠦᠭᠡᠢᠠᠷᠳᠠᠶᠢᠨᠲᠡᠭᠡᠭᠡᠢ	H	A	GEDEG U	USUG BICIᠭᠡᠯᠢ
AYIMAG ALHYJV BOSTIV GYLASARY	ᠠᠶᠢᠮᠠᠭᠠᠯᠬᠢᠶᠢᠪᠣᠰᠲᠢᠪᠠᠶᠢᠰᠠᠷᠢ	R	AYIMAG AL		ORON SAGICA Uᠭ
AYIMAG VSV	ᠠᠶᠢᠮᠠᠭᠠᠯᠪᠠᠶᠢᠰᠠᠷᠢ	Y			NIGE VSV-YI DA
AYIL AYIMAG EYETEI NI SAYIHAN AH_A DEGUU NAYIRTA	ᠠᠶᠢᠯᠠᠶᠢᠮᠠᠭᠠᠯᠡᠶᠡᠲᠡᠢᠨᠢᠰᠠᠶᠢᠬᠠᠨᠠᠬᠠᠳᠡᠭᠤᠨᠠᠶᠢᠷᠲᠠ	G	AYIL-ACA	AYIL-ACA	AYIL-ACA
AYIL-ACA ERHU-BER ASDAR_A-ACA-BAN ERI	ᠠᠶᠢᠯᠠᠴᠠᠢᠷᠬᠤᠪᠡᠷᠠᠰᠳᠠᠷᠠᠴᠠᠪᠠᠨᠡᠷᠢ	G	AYIL-ACA	AYIL-ACA	AYIL-ACA
AYIL-ACA-BAN BVSYD-VN EYODE-YI TADAL UGEI AH_A	ᠠᠶᠢᠯᠠᠴᠠᠢᠪᠠᠨᠪᠦᠰᠢᠳᠦᠨᠡᠶᠣᠳᠡᠶᠢᠲᠠᠳᠠᠯᠤᠭᠡᠢᠠᠬᠠ	H			ᠴᠡᠨᠡᠷᠢᠨᠡᠴᠡᠪᠡᠨ
AYIL ERUHE BVSHARY	ᠠᠶᠢᠯᠡᠷᠦᠬᠡᠪᠦᠰᠢᠬᠠᠷᠢ	Y			GERLEHU
AYIL-IYAN GOLOHV-ACA AGASTI-BAN GOLO AJIL-IYAN GOI	ᠠᠶᠢᠯᠢᠶᠠᠨᠭᠣᠯᠣᠬᠦᠪᠠᠴᠠᠭᠠᠰᠲᠢᠪᠠᠨᠭᠣᠯᠠᠵᠢᠯᠢᠶᠠᠨᠭᠣᠢ	H			AYIL SAGAH_A-T
AYIL-VN ARMAD AYIMAG-VN TURUGU	ᠠᠶᠢᠯᠦᠨᠠᠷᠮᠠᠳᠠᠶᠢᠮᠠᠭᠠᠨᠲᠦᠷᠦᠭᠤ	K			AYIL SAGAHALTA
AYIL-VN ALAG-ACA ADGVN-V ALAG	ᠠᠶᠢᠯᠦᠨᠠᠯᠠᠭᠠᠴᠠᠠᠳᠦᠭᠢᠨᠠᠯᠠᠭ	G	AYIL-VN A		AYIL HOTA-DAGA
AYIL-VN EGUDE AMARAGADAG UGEI	ᠠᠶᠢᠯᠦᠨᠡᠭᠦᠳᠡᠠᠮᠠᠷᠠᠭᠠᠳᠠᠭᠠᠭᠡᠢ	H			AYIL VEHE TOGOᠭ

Fig. 1. The Sample of knowledge base of Mongolian idioms.

²R & D institution of this corpus is Academy of Mongolian Studies, Inner Mongolia University.

³Burintegus compiled, "Mongolian Idioms Dictionary ", Inner Mongolia Education Press, 2001.

TABLE I: FRAMEWORK OF SUB-CATEGORY OF MONGOLIAN IDIOM

Mongolian Idioms(a)					
Phrase of Idioms					
xoršiy_a üge	xebšil üge	xabutu üge	togtamal üge	ner_a tomiy_a	
Y	X	K	J	NT	
Mongolian Idioms(b)					
Sentence of Idioms					
jür üge	sečen üge	xošong üge	ceger üge	surgal üge	lorjoo üge
R	S	H	C	G	L

We have re-defined the various sub-categories and boundaries of the Mongolian idioms from the perspectives of the Mongolian information processing, which will fill up the gap of idiom sub-category marks that have not been set up in the "Information Technology-Information Processing Oriented Mongolian Word and Expression Mark Set" [9]. Thus, a reasonable and integrated Framework of Sub-categories of Mongolian Idiom and its Marks set will be developed. At present, we have set the marks of some sub-categories shown in Table I.

- 4) Topic (VTOROL): fill in the marks of subject classification. For example, fill in "NuM" in the corresponding field of "ᠨᠠᠮᠪᠦᠷᠦᠨᠢᠪᠢᠰᠤᠬᠠᠪᠪᠦᠷᠦᠨᠠᠯᠲᠤ", for the "nomadic topic".

According to semantic and pragmatic properties of Mongolian idioms, we classify the thematic meanings it expressed, and develop a system of "Topic Classification Framework of Idioms" (as follows) and its corresponding marks set. In traditional Mongolian linguistics, the study on idiomatic meanings, the topics that convey cultural themes and cultural connotation was rather weak. So we have limited direct reference materials. When we initially developed the framework of the topic classification, we based on the corpus analysis results, benefited from the previous theoretical studies, we also consulted the recommendations containing cultural connotations in the Mongolian idioms in such works as "A Study on Mongolian Idioms" [10], "The Cultural Heritage of Mongolian Idioms" [11] and so on as the reference.

- 1. Materials (B0d)
 - 1.1.nomadism (NuM)
 - 1.1.1 five kinds of domestic animals (TaM)
 - 1.1.1.1 ... other poultr ies (BvM)
 - 1.2 hunting (AA)
 - 1.3.food (IV)
 - 1.3.1.dairy products (CaI)
 - 1.3.1.1 ...
 - 1.3.2 meat(MaI)
 - 1.4. Residence (OR)
 - 1.5 clothing and accessories (HvJ)
- 2.Wisdom (VV)
 - 2.1 Morals (Y0S)
 - 2.1.1 marriage (H0N)
 - 2.1.2 entertainment (NaN)
 - 2.2 family (GeB)
 - 2.3 faith, sacrifice (\$aT)

be functioned for the parallel idioms so as to offer a convenient tool to promote user's mastery of Mongolian idioms. For an idiom "xala ni gadan_a-ban, xair_a ni dotor_a-ban), for example, the first half of the sentence-"xala ni gadan_a-ban" is shown on the learning interface. The users are to fill the second half of the sentence in the brackets (). If the user does not know how to fill the blanks, he or she can click "next part" to get the entire idiom.

III. MATH THE ANALYSIS OF THE PROSPECT FOR APPLICATION

A. The Explanation of Attributes Field of General Base

A new era of informationization, modernization constantly promote the development of language education, which bring about significant changes in many aspects, in particular of language use. The great changes have also caused the declining trend of idiom use. The serious impacts on the idioms even cause the obvious shrink of using idioms in certain places. Such phenomenon has never happened before. Because of the shrinking vocabulary of Mongolian idioms, resulting in the disappearing cultural connotation that the idioms conveyed, and thus a serious challenge shows up. Facing the predicament of rhetorical function of Mongolian language, and in order to develop and utilize the abundant treasures of the Mongolian language, it is an urgent task for each Mongolian researcher to collect and record the idioms which are being forgotten or lost in the national cultural heritage. This is the best way to repair the Mongolian vocabulary, rescue endangered Mongolian idioms. "In this way, we can ensure the healthy, sustainable development and long-term usage of language resources" [12].

B. The Application in the Fields of Mongolian Information Processing

- 1) Making the computer recognize and understand the knowledge of Mongolian idioms, improving the accuracy of Machine Translation.

The knowledge base will provides formalized description of idiom knowledge for a variety of applications of Mongolian information processing systems. Since Mongolian language belongs to agglutinative language, an important vocabulary feature of it is that a "root word" can not only derive a word and but also contribute to a fixed phrase and even a sentence-formed idiom. For example, as a root word, "nidü-" (eye) can form a verb like "nidü/le/xü" (gaze at) or "nidüle/gde/xü" (to be noted) through suffixes, by means of derivation. In succession, it can form some fixed phrases such as "nidü aldam_a" (boundless), or "nidü üjügürlexü" (to discriminate against), or even constitute the sentence-formed idioms such as "nidün-degen xabçixu ügey" (to look down upon), "nidün-dü orogsan xog,sid ün-dü orogsan mix_a" (pain in the neck) and so on. No matter for a computer or a human being, when meets an idiom, the biggest problem is word meaning, if grammar structure analysis is not in consideration. The knowledge base uses the equivalent or similar word to describe each interpretation of the idiom. As a lexical unit, an idiom indicates a lexical meaning regardless of

its length. This is a new "interpretation" method for the semantic unit of idioms. For instance, the meaning of "xala ni gadan_a-ban" is equal to an adjective "xair_a" (skilled), the meaning of "xair_a ni dotor_a-ban" is equal to a verb "dotor_a" (pride). The new method makes the computer easier to understand and process the idiom, which can solve the problem of semantic processing of the idiom in syntactic processing and Machine Translation, can increase the accuracy of Machine Translation. It can also provide knowledge support for the construction of tree base, semantic network , and what's more, it may provide knowledge support for high-level research such as metaphor, cognitive model.

- 2) Provide a basis of lexical resources for corpus annotation and idiom tagging, lay the foundation for the multilevel processing of Mongolian corpus.

The idiom is a common phenomenon in natural languages. From the analysis of language, the corpus processing is hard to avoid the problem. The research on Chinese idioms and the relevant corpus studies have made remarkable progress. Such as a study of idiom unit based on corpus [13], the mark of idioms [14], research on tagging English and Chinese idioms [15], even its semantic processing model [16], explanatory model of semantic structure [17], etc.. However, formalized researches on Mongolian idioms are relatively underdeveloped. As mentioned above, the base was developed *Dictionary of Modern Mongolian Fixed Phrases with Detailed Description of Grammatical Information (2005)* including more than 7000 commonly used Mongolian fixed phrases, setting 17 fields to describe the grammatical attribute of each fixed phrase. The fixed phrases are defined in the dictionary involving 5 main categories, such as xoršiy_a üge (xoršiy_a üge), xebšil üge (xebšil üge), xabutu üge (xabutu üge), togtamal üge (togtamal üge) and ner_e tomiy_a (ner_e tomiy_a). Compounds are classified into compound nouns, compound adjectives, compound pronouns, compound verbs, and etc. [18]. As the dictionary is unable to cover sentence-formed idioms, we can't achieve the tagging work of sentence form idioms on corpus. As shown in the following example, the underlined part on corpus is not marked.

Example 1:

<</Wp1 TANI/Ve1+LCA/Fe4+HV/Ft12-ACA/Fc41 EM UN_E/OaTAL_A/Ne2-YIN/Fc11 GVRBA/Mu+N/Zx GOR OGESU/Ne1_/Wp1TANI/Ve1+LCA/Fe4+GSAN/Ft11-V/Fc 12H0YIN_A/On-BAN/Fx11TANGNAI/Ne2-YIN/Fc11GV RBA/Mu+N/Zx SVDASV/Ne1>>/Wp1.

We will utilize the "Knowledge Base of Idioms" to tag information of subcategory for the above corpus, as shown in the following example 2 (Here, we use square brackets for idioms, we use slash marks followed by subcategory, "R" represents proverbs.)

Example 2:

<</Wp1[TANI/Ve1+LCA/Fe4+HV/Ft12-ACA/Fc41 EM UN_E/OaTAL_A/Ne2-YIN/Fc11 GVRBA/Mu+N/Zx GOR OGESU/Ne1_/Wp1TANI/Ve1+LCA/Fe4+GSAN/Ft11-V/Fc 1 H0YIN_A/On-BAN/Fx11 TANGNAI/Ne2-YIN/Fc11 GVRBA/Mu+N/Zx SVDASV/Ne1]/R>>/Wp1.

Therefore, the "knowledge Base of idioms" provides basic

resources for constructing a multi-level idiom processing corpus with subcategory tagging and topic tagging.

C. The Value on Language Teaching

On the one hand, the construction of knowledge base of idioms and its software will provide idiom learning platform for Mongolian language education, such as learning mother tongue and writing of primary and middle students, foster and improve the quality of students' mother tongue usage.

On the other hand, it is a new teaching aid service in language teaching for foreigners. Mongolian idiom is a kind of fixed form, which has a rhetorical function, it can produce strong language effects. Therefore, it is an indispensable element for writing and translation. While, in recent years the using of Mongolian idioms has become rare among young people, there even appears the fading phenomenon of Mongolian idioms in many Mongolian teenagers, especially the urban primary and secondary school students. We think that long sentence-formed idioms and parallel-formed idioms are hard part in learning Mongolian idioms. For instance, when students encounter a sentence-formed idiom, such as "ᠠᠨᠠᠭᠤ ᠰᠤᠶ᠋ᠠᠨᠠᠭᠤ ᠰᠤᠶ᠋ᠠᠨᠠᠭᠤ ᠰᠤᠶ᠋ᠠᠨᠠᠭᠤ", the important steps are: the first step is to know the proper segments in terms of its emotive (appreciative, pejorative) values, and to understand the idiom meaning it contains correctly; the second step is the practical application, that is, to apply the idiom to an appropriate occasion or situation. For a parallel formed idiom such as "ᠠᠨᠠᠭᠤ ᠰᠤᠶ᠋ᠠᠨᠠᠭᠤ ᠰᠤᠶ᠋ᠠᠨᠠᠭᠤ ᠰᠤᠶ᠋ᠠᠨᠠᠭᠤ", the key is the correct connection of the first and second parts, preventing deviation in mastery.

IV. CONCLUSION

In Mongolian vocabulary system, there are copious stocks of idioms, and in many categories. In traditional Mongolian research, there remain the researches under discussion, the unclear divisions of certain sub-categories of idioms, uncertain distinctions of some synonymy and variants, and other inconsistencies or deficiencies. Those have been the obstacles for this study. In Mongolian language study, to be particular in language information processing, the generalization and formalized research on semantic features of the idioms are at a beginning stage, such as topic classification according to semantics and pragmatics, and formalized description and analysis on phrase-formed idioms and sentence-formed idioms as a whole. So, this study aims at construct an idioms knowledge base which is helpful for people to understand idioms and contributes to language information processing applications, making full use of previous research achievements, and drawing on relevant theories and methods in English, Chinese, as the reference.

REFERENCES

- [1] Nasun-Urt, "The Presumption of Mongolian Language Resource Platform Framework," in *Collection of Chinese Language Resources Essays*, P. Zhang, T. Wang Eds. BJ: the Commercial Press, 2009, pp. 236-248.
- [2] H. Y. Hua, Nasun-Urt, and Wuyungaowa, "The initial framework of developing semantic knowledge base of mongolian idioms," *Altaic Hakpo*, vol. 22, pp. 121-139, June2012.
- [3] G. Akim, *Brief dictionary of mongolian idioms*, 1st ed. Hohhot, CH.: Inner Mongolia Cultural Press, 2000.
- [4] B. Comp, *Dictionary of Mongolian Idiom*, 1st ed. Hohhot, CH.: Inner Mongolia Educational Press, 2001.
- [5] D. Qinggeltu, *Dictionary of Modern Mongolian Fixed Phrases with Detailed Description of Grammatical Information*, 1st ed. Hohhot, CH.: Inner Mongolia educational press, 2005, pp. 3-5.
- [6] D. Qinggeltu, *A study on compound words for Mongolian Information Processing*, 1st ed., Hohhot, CH.: Inner Mongolia educational press, 2001.
- [7] C. Norjin et al., *Dictionary of the Mongolian Language*, 1st ed. Hohhot, CH.: Inner Mongolia people press, 1997.
- [8] Mongolian Academy of Science, *Mongolian Dictionary with Detailed Explanation*, 1st ed. Ulaanbaatar, 2008.
- [9] *National Standard of the People's Republic of China (GB/T26235 2010)*, Information Technology-Information Processing Oriented Mongolian Word and Expression Marks Set, the state administration of quality supervision, inspection and quarantine of the People's Republic of China, 2010.
- [10] H. Wang, "A Study on Mongolian idioms," Ph.D. dissertation, Dept. Elect. Eng., Inner Mongolia Univ., Huhhot, pp. 113-128, 2010.
- [11] Y. Chen, "The cultural heritage of Mongolian idioms," M.S. thesis, Dept. Electron. Eng., Inner Mongolia Normal Univ., Hohhot, China, pp. 18-61, 2010.
- [12] Z. Chen, "On language resources," in *Collection of Chinese Language Resources Essays*, P. Zhang, T. Wang, Eds., BJ: the Commercial Press, 2009, p. 13.
- [13] J. Yang, "Dynamic circulating corpus (DCC)-based Idiom unit of Chinese," Ph.D. dissertation, Dept. Elect. Eng., Beijing Language Univ., Beijing, 2005.
- [14] N. An, H. Liu, and M. Hou, "Tagging of the Idiom in the corpus," *Journal of Chinese Information Processing*, vol. 18, pp. 18-25, January 2004.
- [15] N. An, "Research on tagging the English and Chinese Idiom Based on Corpus," in *Proc. the Second Student Workshop on Computational Linguistics (SWCL2004)*, pp. 80-85, August, 2004.
- [16] F. Zhu and H. Zhang, "The semantic processing model and its influencing factors of idioms," *Foreign Languages Research*, vol. 104, April 2007.
- [17] H. Zhang and F. Li, "Discussion on the explanatory model of semantic structure of idioms," *Foreign Languages and Their Teaching*, vol. 234, September, 2008.



Hai Yinhua was born in Inner Mongolia, China on September 27, 1981. She received her bachelor degree of language and literature from September 1999 to June 2003, Inner Mongolia Normal University (the major is Mongolian language and literature), Hohhot, Inner Mongolia, China, master of literature from September 2003 to June 2006, Inner Mongolia Normal University (the major is Mongolian language information processing), Hohhot, Inner Mongolia, China, and doctor of literature, from September 2006 to June 2010, Inner Mongolia University (the major is Mongolian language information processing), Hohhot, Inner Mongolia, China.

She is a doctor of School of Mongolian Studies, Inner Mongolia University. She is teaching natural language processing since September, 2010. She has been published about 30 papers on domestic and foreign journals in the research of Mongolian information processing.

Dr. Hai is a membership of Chinese Information Processing Society of China.