Ontology Method for Chinese Language Processing

Longwei Qian, Mikhail Sadouski, Wenzu Li
Department of Intelligent Information Technology
Belarusian State University of Informatics and Radioelectronics
Minsk, Belarus
qianlw1226@gmail.com, mishachess@mail.ru, wzzggml@gmail.com

***Abstract*—This paper describes a method for computer processing of Chinese language texts based on Chinese linguistic ontology. The ontology is used to build a unified semantic model of Chinese linguistic knowledge and effectively organize the domain knowledge for processing of Chinese language texts. The described method of texts processing could be used further for Chinese language user interface implementation.**

***Keywords*—Chinese language processing, ontology, knowledge-driven, Chinese language generation**

I. Introduction

Natural language processing has always been one of the core contents of research in the field of artificial intelligence. It is also the key basic technology for implementing natural language user interfaces. Natural language user interface understands users' intent through learning and understanding natural language, then uses the representation form of natural language to serve users through the interface.

A. Problems of Chinese Language Processing

Due to the diversity and openness of natural language, the computer processing (understanding) of different natural languages has own characteristics and special difficulties. Chinese language processing (CLP) is an exclusive field dedicated to conducting relevant research on Chinese languages and characters [1]. Researches on Chinese language processing show that machine translation systems from foreign languages (especially English) to Chinese language are far superior in terms of performance compared to machine translation systems from Chinese language to foreign languages. Even if application systems such as information retrieval and information extraction, which do not require deep syntax analysis, must overcome the obstacles of word segmentation and part-of-speech tagging in Chinese language processing [2].

Compared with other natural language processing technologies, the reasons that make Chinese language processing difficult can mainly be divided into the following aspects:

- the same part-of-speech can serve multiple syntactic components without morphological changes, i.e. in Chinese language texts regardless of any syntactic component served by each part-of-speech its morphology does not change;
- the construction principles of Chinese sentences are basically consistent with the construction principles of phrases. According to the different composition method, the structure of Chinese phrases can be divided into various types such as subject-predicate, verb-object, modifier-noun, joint and so on. However, if Chinese sentences is directly composed of words, the structure of Chinese sentences are also basically these;
- the function of functional words in Chinese language is complicated. The function words generally refer to words that do not have a full meaning but have grammatical meaning or function. Due to the lack of morphological changes in Chinese language texts, sentence patterns, and even tenses in sentences need to be represented by functional words;
- the writing habits in Chinese language. European languages such as English, Russian are basically written in words, with natural spaces between words. However, in Chinese language text there is no natural space to separate each word.

B. Analysis of Existing Approaches for Chinese Language Processing

The current natural language processing methods are divided into the following directions [3]:

- the methods of establishing logical reasoning system based on rules and knowledge base and based on Noam Chomsky's grammar theory;
- the machine learning methods based on large-scale corpus and based on mathematical statistics and information theory.

The statistical methods avoid the error situation caused by the subjective factors of manual writing rules in the rule-based methods, and reduce the dependence of the system on human. However, it inevitably has the defects

that the performance of the model is too dependent on training samples, and it lacks domain adaptive capabilities.

At present, whether the rule-based methods or the statistical-based methods, the methods and evaluation standards used in Chinese language processing are almost borrowed from European language processing methods such as English. The analysis process lacks the characteristics of Chinese language texts [4].

C. Analysis of Existing Approaches for Chinese Language Generation

After the natural language interface analyses the natural language, the intelligent system needs to process these information, then to generate natural language texts and display it for users. This paper studies the process of generating the natural language texts from the semantic structure in the knowledge base. The semantic structure is constructed using a formal semantic network language. In the field of natural language processing, there are many researches on analyzing natural language texts to transform them into semantic networks. But there are few researches on generating natural language from formal semantic network structure.

The traditional natural language generation process is to transform a tree structure to natural language texts. However, the information structure formed by the semantic network language is a network topology. Simmons discoursed the use of augmented transition networks to generate English sentences [5]. Shapiro used extended transformation network syntax to implement sentences generation from semantic network [6]. These methods use fixed templates for sentence generation and sentence planning, which lack flexibility and adaptability. Dai Yintang studied the semantic network language generation, transformed a complex semantic network into multiple semantic trees, then serialized into language nodes [7]. This method improves the flexibility of sentence generation, but lacks the constraints of language knowledge, so the generated language fluency is insufficient.

D. Analysis of Existing Knowledge Base for Chinese Language Processing

Faced with the problems of Chinese language processing, Liu Zhiyuan from Tsinghua university proposed adding knowledge to the data-driven natural language processing model, and researched the natural language processing model driven by knowledge and data [8]. The knowledge of human can be divided into domain knowledge, common sense knowledge and linguistic knowledge. These complex knowledge help intelligent systems to achieve the deep understanding of natural language texts. The various knowledge can reasonably explain natural language processing. The linguistic knowledge can provide basic features and basic data resources for algorithms. The well-known knowledge bases for the natural language processing such as Freebase, DBpedia and so on, in the file of Chinese language processing, there are CN-DBpedia [9], zhishi.me, et al.

So far in the field of Chinese language processing, many excellent **language knowledge bases** have been developed.

The "Grammatical Knowledge-base of Contemporary Chinese (GKB)" [10], developed by Institute of Computational Languages of Peking University. It contains 73,000 Chinese words in accordance with the criteria of combining grammatical function and meaning. In accordance with the principle of grammatical function distribution, the various grammatical attributes of each word are classified and described.

The Chinese FrameNet ontology [11], developed by Shanxi University, the Chinese FrameNet ontology is constructed by drawing on the ontological thought of the Framenet project and its strong semantic analysis capabilities. The semantic frame is used to identify the relationships between lexis, and a valence mode is provided to reflect the syntactic-level relationships. So that the computer can perform a deeper semantic analysis of natural language texts.

The "HowNet" [12], developed by the professor Dong Zhengqiang. "HowNet" considers the lexis or Chinese characters is the smallest linguistic unit, but not the smallest semantic unit. It considers that the lexis can be described in a smaller semantic unit - sememe.

The GKB is organized according to the database model, and knowledge does not constitute semantic relationships. The theoretical basis of Chinese FrameNet Ontology is frame semantics, which faces narrow lexis coverage rate and insufficient description of syntactic information. The "HowNet" focuses on the description of semantic of Chinese lexis and concepts, lacks the description of grammatical and semantic information of the Chinese language texts.

E. The Proposed Approach

This paper takes a basis of these excellent Chinese language knowledge bases, proposes to construct the language ontology for Chinese language processing based on open semantic technology for intelligent systems (OSTIS technology). The OSTIS technology is aimed for presentation and processing of various knowledge, it's focused on the development of knowledge-driven computer systems [13]. The ontological approach of OSTIS for Chinese language processing is to construct a hierarchical system of subject domains of Chinese linguistic and their respective ontologies.

The Chinese linguistic knowledge could be described from various levels of Chinese linguistics such as lexis, syntax and semantics. This method uses the ontological

thought to reasonably organize the knowledge and structure the knowledge in the file of Chinese linguistics. Based on OSTIS technology, the subject domains and respective ontologies of various levels of Chinese linguistics are constructed for Chinese language processing.

With the help of different levels of Chinese linguistic knowledge, it is beneficial to ambiguity resolution in Chinese language processing and content selection in Chinese language generation. In the process of machine translation, Chinese linguistic ontology can be used as a conceptual source for intermediate languages. The knowledge-driven Chinese language processing model can effectively improve the performance of Chinese language processing.

II. The structure of the knowledge base of the Chinese language interface

The knowledge base of natural language processing developed based on OSTIS technology is a part of the common linguistic knowledge base for all ostis-systems. Based on the general structure of linguistic subject domains, the development of Chinese linguistic subject domains and specific knowledge processing is the proposed method in this paper to implement the Chinese language interface that could process Chinese language texts.

According to the principle of using the OSTIS technology to construct the knowledge base, the construction of the knowledge for Chinese linguistics is a hierarchical system of the various subject domains (SD) of the Chinese linguistic knowledge and their corresponding ontology. The knowledge formalization can be considered as a formalization and specification of a subject domain, including a structural specification of a subject domain, terminological ontology, theory-set ontology, logical ontology and so on [14].

Based on OSTIS technology Chinese linguistic ontology classifies knowledge according to the subject domains. The various subject domains contain the different aspects of corresponding Chinese linguistic knowledge, as well as the rules and methods for operating the knowledge.

For Chinese language the grammatical features are not very obvious. A reasonable description of lexical semantic knowledge is more important for Chinese language processing. In addition to the previously described knowledge base. there are excellent lexis knowledge bases such as WordNet, VerbNet, et al. They have important reference significance for constructing lexical ontology. The domain ontology of Chinese linguistics is constructed by reusing existing ontologies and linguistic knowledge oriented to Chinese language processing. In the construction process of linguistic ontology, drawing on the reuse of existing Chinese linguistic ontology can improve the efficiency of ontology construction.

In this paper the process of constructing the domain ontology of Chinese linguistics can be roughly divided into the following 4 steps:

- determine that the requirements of the constructed ontology are oriented to Chinese language processing;
- research domain knowledge and reusable ontology;
- use specific examples to clarify subject domains, specific concepts and various relationships between these concepts;
- evaluation and improvement of ontology.

The goal of the proposed method is to construct syntactic ontology and semantic ontology for Chinese language processing using the bottom-up approach of domain ontology construction.

The development of subject domain of Chinese linguistic and their corresponding ontology using OSTIS technology uses semantic networks representation - SC-code (Semantic Code). Semantic networks representation is in the form of semantic network language with the basic set-theoretic interpretation.

The following is the general structure of the Chinese language subject domain represented in SCn-language [13].

SD of Chinese language texts
⇒ *particular SD**:
- *SD of Chinese language syntax*
- *SD of Chinese language semantic*

The *Subject domain of Chinese language syntax* describes the characteristics of the Chinese language syntax, the functional characteristics of the syntactic components. The *Subject domain of Chinese language semantic* describes the semantic characteristics of lexis, the semantic relationships and the semantic structure in the Chinese language texts.

Thus the key to the proposed method is the description of these specific subject domain and its corresponding ontology. The key elements of a specific subject domain are the main concepts studied in the subject domain and relationships. The relationships between these specific subject domains also need to be considered in the ontology. In general, the entire Chinese linguistic ontology forms a hierarchical and multilevel semantic model.

The focus of current research is on the subject domain of Chinese syntax, which will be described in detail below.

In the *Subject domain of Chinese language syntax*, the syntactic and semantic information of sentences, phrases and lexis in Chinese language texts need to be considered. These specific subject domains could provide linguistic knowledge for Chinese language processing.

The following is the structural fragment of the subject domain of Chinese language syntactic represented in SCn-language.

SD of Chinese language syntax
⇒ *private SD**:
- ***SD of Chinese sentence pattern***
- ***SD of Chinese language phrase***
- ***SD of Chinese grammatical information of the lexis***
- ***SD of Chinese function words***
- ***SD of non-lexis of Chinese language***

The *SD of Chinese sentence pattern* indicates types of the sentence, syntactic and semantic relationships between sentences. The *SD of Chinese language phrase* indicates types of the phrase, the internal structure and external functions of basic Chinese language phrases, syntactic and semantic relationships between phrases. The *SD of Chinese grammatical information of the lexis* indicates classifications of Chinese language lexis, and their grammatical and semantic functions of lexis. The *SD of Chinese function words* describes the types, functional characteristics of functional words, as well as grammatical attributes of functional words. The *SD of non-lexis of Chinese language* describes the features and syntactic functions of non-lexis in Chinese language texts.

The description of *Subject domain of the non-lexis components* in the Chinese language texts conforms to the characteristics of Chinese hieroglyph, some language components don't conforms to the definition of words in Chinese language. In European languages the language component smaller than word is the letter, but in the Chinese language texts the non-lexis components not only include Chinese characters, but also the idioms and the abbreviations.

Sentences have always been the smallest research unit in the field of natural language processing. Sentence analysis is an important intermediate stage connecting discourse analysis and lexis analysis. The detailed description of sentence knowledge is an important basic stage in the entire natural language processing. The *Subject domain of Chinese sentence pattern* studies the types of sentence patterns in Chinese language, determination of sentence components and their relationships, i.e. the division of sentence components in a simple sentence and an accurate functional description of sentence components.

The following is the structural fragment of the *Subject domain of the sentence pattern* represented in SCn-language.

SD of Chinese sentence pattern
∋ *maximum studied object class'*:
sentence pattern
∋ *not maximum studied object class'*:
- *simple sentence*
- *compound sentence*

∋ *explored relation'*:
- *subject'*
- *object'*
- *head predicate'*
- *attribute'*
- *center word'*
- *adverbial modifier'*
- *complement'*

Simple sentence
⇐ *subdividing**:
{
- *subject-predicate sentence*
- *non-subject-predicate sentence*
- *special sentence*

}

The sentence structure of a simple subject-predicate sentence can be described by the ontology. For example, it can be composed of a noun phrase and a verb phrase according to subject predicate relationship (Fig. 1).

As a highly deterministic partial analysis of Chinese language texts, the analysis of phrases can solve most of the problems of local ambiguity structure. As an intermediate result of sentence analysis, phrase analysis are also the basis of deeper chunk analysis and complete syntax analysis. The *Subject domain of Chinese phrases* studies the types of Chinese phrases and the relations between the internal structure of these phrases.

The following is the structural fragment of the *Subject domain of the phrases* represented in SCn-language:

SD of Chinese phrase
∋ *maximum studied object class'*:
phrase
∋ *not maximum studied object class'*:
- *noun phrase*
- *verb phrase*
- *adjective phrase*
- *numeral classifier phrase*

∋ *explored relation'*:
- *subject'*
- *object'*
- *head predicate'*
- *attribute'*
- *center word'*

Verb phrase
⇐ *subdividing**:
{
- *transitive verb phrase*
- *intransitive verb phrase*

}

Various basic phrases structures can be described by ontology according to syntactic relationships. For example, a verb phrase can be composed of verb and other any a

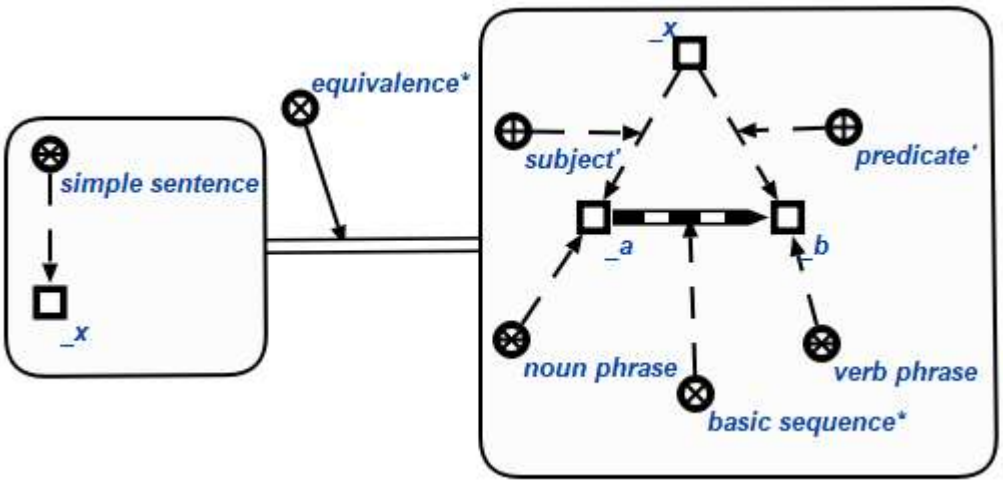

Figure 1: Logical statement about subject predicate sentence

basic phrase or verb and a noun according to predicate-object relationship (Fig. 2).

Through the structure of the ontology of phrase knowledge and the sentence pattern described above, the construction of the ontology of Chinese language knowledge conforms to the structural consistency of phrases and sentences in Chinese language texts.

Due to the problem of Chinese writing habits, the standardization of words in Chinese language processing is the first theoretical problem. In the field of Chinese language processing, the "Contemporary Chinese Language Word Segmentation Standard Used for Information Processing" has been proposed. In this standard, the word are represented by "segmentation unit". The precise definition is: "the basic unit for Chinese language processing with certain semantic or grammatical functions". According to the knowledge of Chinese linguistics, the Subject domain of Chinese lexis studies types, syntactic functions and semantic functions of lexis. The following is the structural fragment of the *Subject domain of the lexical grammatical information* represented in SCn-language:

SD of lexical grammatical information
∋ *maximum studied object class'*:
lexical grammatical information
= the grammatical information of the segmentation unit in Chinese language
∋ *not maximum studied object class'*:
- *noun*
- *verb*
- *adjective*

Verb
⇒ *inclusion**:
- *verb that do not take object*
- *verb that take object*
- *verb that take double object*

Based on the ontology method of OSTIS technology, the Chinese linguistic knowledge is structured for processing of Chinese language texts. The goal of constructing knowledge base based on ontology is to provide a common understanding of Chinese linguistic knowledge, determine commonly recognized concepts. By constructing the knowledge base of Chinese linguistic knowledge, the clear definition of the concepts and their interrelationships in Chinese linguistic knowledge is determined. The ontology also contains the main theories and basic principles of Chinese language, as well as the methods and rules of knowledge processing. The knowledge base divides Chinese linguistic knowledge into various subject domain, it's beneficial to the management and application of knowledge. Thus the knowledge base of Chinese language has the ability for Chinese language processing. The Chinese language interface can use the concepts, semantic relationships, inference rules and other knowledge in the knowledge base to achieve Chinese language processing.

III. Implementation of Chinese Language Processing

Automatic analysis of natural language is a hierarchical process, which is roughly divided into lexical analysis, syntactic analysis and semantic analysis. Generally, there are some stages for the process of the natural language text understanding with the aid of the OSTIS-system [14]. Based on these steps the proposed method uses discourse analysis (context) of the input text. Context analysis is performed at each specific level of natural language automatic analysis and interaction with the content in the internal knowledge base or external information resources during the analysis process. The analysis process of the

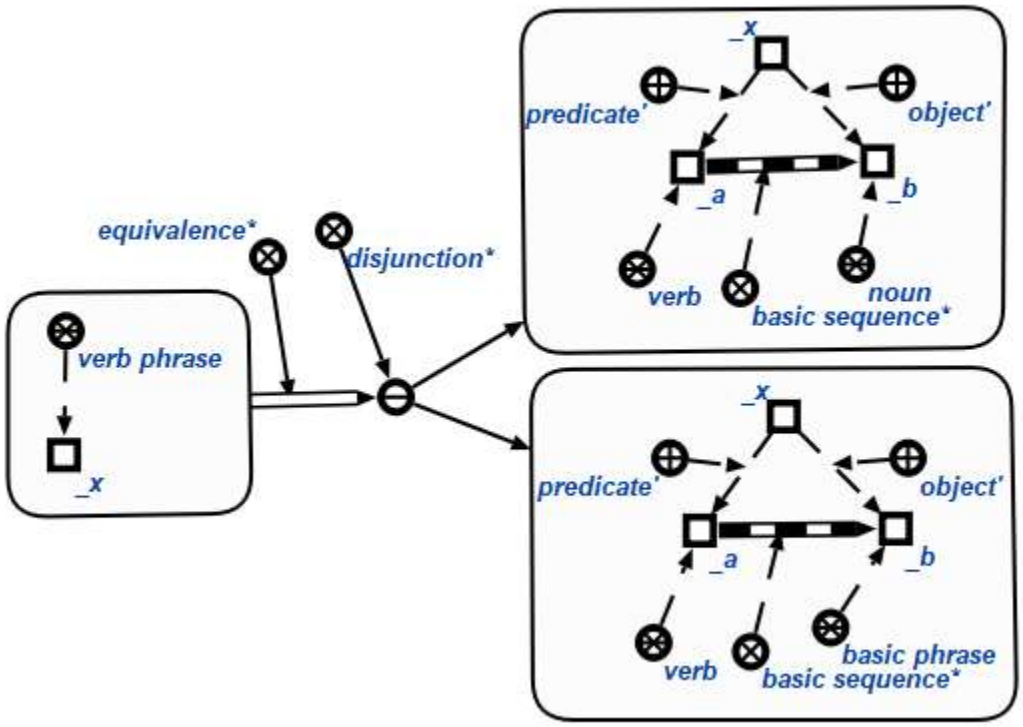

Figure 2: Logical statement of a verb phrase

natural language texts based on OSTIS technology [15] can be roughly divided into the following sections:

- Linguistic (graphematic) analysis generates a series of independent words in a given order;
- Lexical morphological reduction, lexical analysis, context analysis (comparative analysis with the content in the knowledge base of a specific domain);
- Syntactic tree generated, the specific agent transforms it into a semantically equivalent structure in the knowledge base;
- Semantic Analysis: through semantic ontology in linguistics and ontology knowledge in the specific subject domain, the agent correctly analyzes the semantics of the sentence.

The analysis of natural language texts based on OSTIS technology needs to be adjusted specifically to suit the characteristics of Chinese language texts.

Due to the writing habits of Chinese language text, there are no natural spaces between words. So the graphematic analysis need determine the basic text components: segmentation units, e.g. with the help of third-party software [16], an ordered sequence of Chinese characters is segmented into correct independent words with a given order. The lexical context analysis process is combined with the Chinese language texts segmentation process. Comparing the results of Chinese language texts segmentation with the content in the knowledge base of a specific domain, the accuracy of the Chinese language texts segmentation results can be improved. The agent transforms the results of the syntactic analysis into a semantically equivalent fragment in the knowledge base, the structure of fragment in the knowledge base is constructed in the semantic network language. By combining with the knowledge of linguistic ontology and the knowledge base of the specific domain, the information of natural language texts is integrated into the existing knowledge base to realize the deep semantic understanding of Chinese language texts.

The Chinese language processing described in the above section belongs to the study of natural language understanding. Another part of natural language processing is natural language generation. Natural language generation and automatic analysis of natural language are not simple and mutually inverse processes. There are obvious differences in both the implementation model and the basic process. In the field of natural language generation, the classic pipeline model consists of three parts: document (content) planning, sentence planning and surface realization. Corresponding to classic model, the natural language generation process is divided into the following three parts:

- Content selection: Choose a suitable and complete semantic structure, which can represent a complete theme;

- Content unit division: A complete semantic structure is divided into one or more semantic substructures, these substructures is interrelated with semantic relationships. Each semantic substructure is suitably represented by a natural sentence;
- Sentence generation: Transform the semantic substructure into a natural language sentence, then the complete semantic structure is transformed into natural language texts.

The process of selecting content for a concept and presenting it to the interface through various output forms is called semantic presentation. The application of semantic presentation in this paper is to use a unified method for content planning and sentence planning for the related knowledge of a central concept to generate natural language texts. According to the user's intention, the user interface identifies the central concept of solving user problems and determines the boundaries of the semantic structure in the knowledge base. Based on the knowledge of the Chinese language ontology and the content of the specific domain knowledge base, the confirmed semantic structure could be transformed into the highly readable natural language texts on the interface for users.

The quality of Chinese language sentences generated from semantic substructures is the basis for the entire semantic structure to be transformed into Chinese language texts. The specific process of generating Chinese language sentence from semantic substructure is divided into: determining the grammatical information of lexis, phrase generation and sentence generation.

The grammatical information of the lexis is determined by the constructed ontology (Fig. 3), i.e. the syntactic functions of the lexis itself and the ability to combine with other notional words and function words and so on.

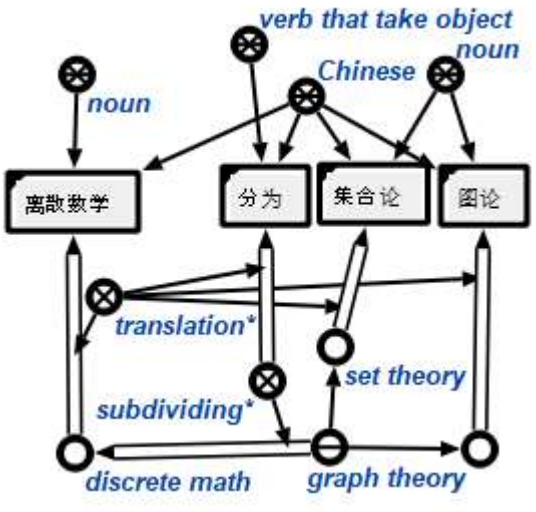

Figure 3: Grammatical information of lexis

Lexis constitutes phrases or sentences through their syntactic and semantic relations. According to the principles of Chinese linguistics, in the process of Chinese language generation, firstly lexis constitute the smaller language units — phrases. The syntactic and semantic function of the lexis in the phrase are determined by the phrase knowledge ontology, then the lexis constitute the phrase (Fig. 4).

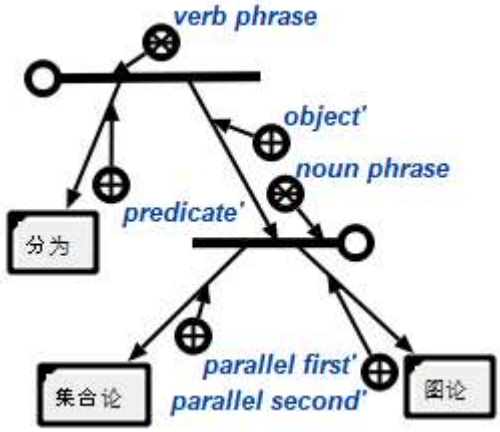

Figure 4: Phrase generation

In the Figure 4 the noun phrase consists of two nouns according to parallel relation. The noun phrase is generated using the rules of Chinese linguistics described in the logical ontology. In the Figure 4 the noun phrase could be composed by the following rule: noun plus conjunction, then plus another noun. The verb phrase are composed of the verb and the noun phrase according to the predicate-object relation. The verb serves as predicate in the verb phrase, and the noun phrase serves as object.

The ontology of sentence pattern knowledge confirms the syntactic and semantic functions that the phrases serve in a sentence, and then phrases "achieve" the sentence (Fig. 5).

Structured knowledge in the knowledge base of Chinese language is applied in the process of Chinese language processing. In the process of Chinese language understanding, the knowledge in the knowledge base of Chinese linguistic provides analysis of lexis, phrases, and sentences in Chinese language texts. The Chinese language texts could be transformed into semantically equivalent structures of specific domain. The semantically equivalent structure is fused with the knowledge in knowledge base of specific domain to achieve the deep understanding of Chinese language texts. In the process of natural language generation, the knowledge in the knowledge base of Chinese linguistics is used to determine the syntactic and semantic function of Chinese language for each semantic node in the semantic structure. The knowledge of Chinese linguistics is used to explain each part of the language generation process, making the

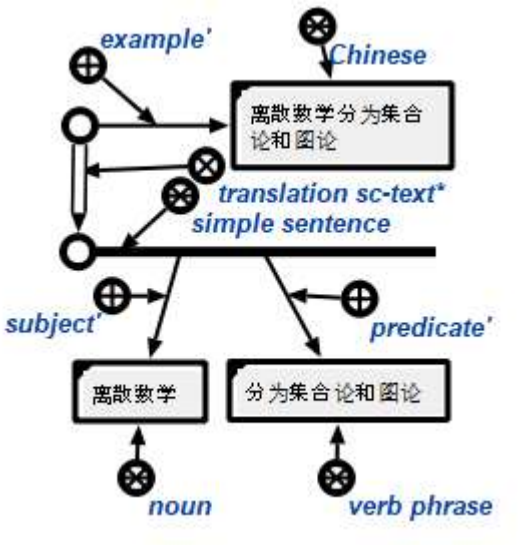

Figure 5: Sentence generation

generated Chinese language texts more accurate. Thus the Chinese language interface is implemented using the Chinese language processing method designed based on ontology.

IV. Conclusion

The proposed method of Chinese language processing is intended to developing the semantic model of Chinese language knowledge. The semantic model can be implemented on different platforms without changing the model itself. In the process of Chinese language automatic analysis, the domain knowledge base can be used as a word segmentation dictionary to achieve more accurate Chinese word segmentation. Based on the ontology of Chinese language knowledge Chinese language processing can be explained with reasonable linguistic knowledge. Obviously, the proposed method in this paper is the preliminary result of the research work of Chinese language processing based on ontology. The improvement of Chinese syntactic ontology, the development of Chinese semantic ontology is the focus of the next work. The more comprehensive context analysis in Chinese language automatic analysis, the problem of the semantic structure planning in Chinese language generation need be more in-depth researched.

References

[1] Zong C.Q. Chinese Language Processing: Achievements and Problems. Chinese Journal of Language Policy and Planning, 2016, vol. 01, No 06, pp. 19-26. (in Chinese)

[2] Yu S.W. Zhu X.F. Wang H. New Progress of the Grammatical Knowledge-base of Contemporary Chinese. Journal of Chinese Information Processing, 2001, vol. 15, No 01, pp. 59-65. (in Chinese)

[3] 2018 Natural Language Processing Research Report. Available at: https://aminer.org/ (in Chinese)

[4] Zong C.Q. Cao Y.Q. Yu S.W. Sixty Years of Chinese Information Processing. Applied Linguistics, 2009, vol. 01, No 04, pp. 53-61. (in Chinese)

[5] R. F. Simmons Generating English discourse from semantic networks. Communications of the ACM, 1972, vol. 15, No. 10, pp. 891-905.

[6] S. C. Shapiro, Generalized Augmented Transition Network Grammars for Generation from Semantic Network. American Journal of Computational Linguistics, 1982, vol. 08, No 01, pp. 12-25.

[7] Dai Y.T. Research on the Key Techniques of Semantic Networks based on Knowledge Collaboration. CA: Fudan University, 2009. (in Chinese)

[8] Liu Z.Y. Knowledge guided natural language understanding. China Conference on Knowledge Graph and Semantic Computing, 2018, pp. 199-206. (in Chinese)

[9] Xu B. Xiao Y.H. CN-DBpedia: A Never-Ending Chinese Knowledge Extraction System. In International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems, pp. 428-438. Springer, Cham, 2017.

[10] Yu S.W. Zhu X.F. Modern Chinese Grammar Information Dictionary, Peking University Open Research Data Platform. [online]. Available at: http://dx.doi.org/10.18170/DVN/EDQWIL.(in Chinese)

[11] Jia J.Z. Dong G. The Study on Integration of CFN and VerbNet, WordNet. New Technology of Library and Information Service, 2008, No 06, pp. 06-10. (in Chinese)

[12] Hownet knowledge database. Available at: http://www.keenage.com/. (in Chinese)

[13] Golenkov V.V. Gulyakina N.A. Proekt otkrytoi semanticheskoi tekhnologii komponentnogo proektirovaniya intellektual'nykh sistem. Chast' 1 Printsipy sozdaniya [Project of open semantic technology of component designing of intelligent systems. Part 1 Principles of creation]. Ontologiya proektirovaniya [Ontology of designing], 2014, no 1, pp. 42-64. (in Russian)

[14] Maksimov V.Yu., Klyshinsky ES. The problem of understanding in artificial intelligence systems. New information technologies in automated systems, 2016, pp. 43-60

[15] Hubarevich N. Boyko I. Knowledge acquisition based on natural language texts. Open semantic technologies for intelligent systems, 2018, pp. 199-206

[16] Language Technology Platform. Available at: http://www.ltp-cloud.com/

Онтологический подход к обработке китайского языка

Цянь Лунвэй, Садовский М.Е., Ли Вэньцзу

В статье рассматриваются существующие методы обработки китайского языка, а также методы построения онтологии предметной области. Был проведен анализ проблем, возникающих при обработке китайского языка в настоящее время.

На основании различных рассмотренных методов построения баз знаний английского и китайского языка был предложен онтологический подход к обработке китайского языка. Предложенный метод направлен на разработку семантической модели знаний о китайском языке. Как один из этапов реализации подхода была создана онтология китайского языка, которую можно использовать в дальнейших этапах для обработки китайского языка.

Таким образом, для дальнейшей обработки и реализации естественно-языкового пользовательского интерфейса в работе предлагается модель обработки китайского языка, основанная на знаниях.

Received 10.01.2020