International Journal of Multimedia and Ubiquitous Engineering
Vol.9, No.6 (2014), pp.209-218
http://dx.doi.org/10.14257/ijmue.2014.9.6.21

English Semantic Feature Processing and Sentence Structure Analysis Based on Hierarchical Network of Concepts

Bangqing Pei

School of Foreign Languages, Mianyang Normal University, Mianyang, China, peibangqing@163.com

Abstract

Current English-Chinese translation machine can't understand the source sentence fully. The theory of hierarchical network of concepts is proposed for resolving the problem through methods including function, translator, effect, relation and state. English semantic feature plays an important role in analysis of sentence comprehension. In our study, we discuss the core structure of English semantic respectively and put forward the corresponding of source sentence compared with online machine translation machines .The experimental results can provide technical support for English -Chinese machine translation. Effect chain and judge is called generalized function effect chain, it is not only the basis of element concept classification, but also the foundation of sentence semantic classification.

Keywords: *English semantic feature, NHC, translation, natural language processing*

1. Introduction

Hierarchical Network of Concepts put forward an effect chain thought which can reflect commons among things, descript the existence and development of things based on basic discipline [1]. HNC is short for Hierarchical Network of Concepts. It is a theory about the natural language understanding and processing. HNC is called the theory of hierarchical network of concepts which is based on basic conceptualize, hierarchical and network semantic expression. HNC theory divided human cognitive structure into local and global associative network. The theory considered the express of associative network is a fundamental problem of expression language.

HNC theory is the sentence semantic types, including 57 basics sentence categories, 3192 groups of mixed sentences and over 10 million groups of compound sentence category [2]. Basic sentence is a sentence category used for describing generalized function effect chain. The sentence may have one Eigen Chunk (EK) or even no EK , Mixed sentence category consists of basic sentence blend , is described generalized function effect chain with two or even links in the sentence and mixed sentence, the sentence in the presence of two or more EK, which contains different generalized function effect chain information.57 basic sentences can be divided into action sentence, effect sentence ,process sentence, relationship sentence, status sentence, judgment sentence and status sentence.

English is a structured, logical language and the predicate center for subject-verb mechanism is very prominent [3]. Complex English sentence contains more than one subject predicate structure. The main elements of complex sentence are formed of prepositional phrase and participle phrases attribute, adverbial, independent component, and a fixed expression. The elements introduced by insertion of constituents and conjunction compound sentence and relative pronouns, adverbs cited. The analysis of complex sentence structure is an important factor for the quality of translation [4].Machine translation research is aim at

ISSN: 1975-0080 IJMUE

study language understanding and generation, while premise fundamental of generate is understand [5-6].The current machine translation software usually analysis sources statement and no depth to the semantic level, which resulting in the accuracy of current machine translation. The understanding of language thought based on HNC theory combines syntactic with semantic layer face for language comprehension. In our study , analysis of English semantic feature block structure based on the HNC theory was carried out , we proposed an English semantic feature block computer processing algorithm provides deeply understand of sentences and better translation of the meanings.

2. Related theory of Hierarchical Network of Concepts

2.1. Translation Principle of HNC Machine

HNC theory puts forward a new thought for machine translation which is viewed as a complex mapping. HNC can be divided into the following three mapping: G1: Sentence Group of Source Language >Sentence Group of Source; G2: Sentence Group of Source Sentence Group of Target; G3: Sentence Group of Target> Sentence Group of Target Language.

G1 is mapping derived from the source language concept space, it corresponds over the understanding process translation system. The sentence group of source language (SGSL) is mapped into sentence group of source (SGS). G2 is a mapping from the source language to target language based on concept space of language. It corresponds to the translation process of translation system. The sentence group of source (SGS) is mapped into sentence group of target (SGT). G3 is a mapping from the concept space to target language space. It is the process of language. Sentence group of target (SGT) is mapped into sentence group of target language (SGTL). The three mapping relied on a form of language concept space. The computer can process natural language completely through several primitives of language concept space such as concept, sentence category and context unit.

2.2. Expression Patterns of HNC

In HNC theory, concept is in infinite while concept elements are finite, finite concept can be expressed by finite concept elements. Basic unit concept, basic concept and logic concept are three basic concepts of HNC design abstraction. They post the primitive and system of abstract concept. The semantic web is treelike hierarchical structure, each layer of the plurality of nodes are expressed numerically. Every node of the network can start from the tops and determined by unique number, the digit string is called the concept of the HNC symbol. The basic concepts, basic unit concept and logic concept are the three conceptual categories of HNC theory. The infinite concept natural language is described through these three kinds of concept. The logic of the concept usually relate to corresponding language words such as prepositions and conjunctions. Design concept of logic is aimed to establish variety marks of semantic chunk. They served the sentence category analysis of the semantic chunk perception. The concept of diversity in natural language expressed as speech phenomena. The HNC theory describes the abstract concept from dynamic (v), static (g), property (u), values (z) and effects (r). If a word is from one side to express a concept, it will be known as one of the five concepts. Concepts are related, such as “student” and “school”, “car” and “road” HNC calculate the association between concepts through the concept of correlation function [7-11]. Sentence semantic chunk is represented as the formula (1).

$$\mathrm{FJ} = \sum_{i=0}^{m} \left(JK_i \right) + \mathrm{E} + \sum_{j=0}^{m} \left(JK_i \right) \quad (1)$$

FJ represent the while sentence, JK represent a generalized object semantic chunk. The sentence is the sentence semantic type. In HNC theory, sentence is infinite while sentence elements are finite, infinite sentence can be expressed by finite sentence elements. The standard of sentence category dividing is called effect chain + judgment. Different sentence type has its own characteristics, such as semantic block number and type, the semantic block combination modes, which are called the sentence category knowledge.

HNC theoretical study of these basic sentence type of sentence category knowledge, establish the sentence category knowledge base, in order to understand sentences. Semantic chunk is the component sentence semantic and the lower level unit of sentence. Semantic blocks have main and auxiliary branch. Subject sense block is the sentential semantic necessary trunk, equivalent to a grammar of subject-verb-object sentence semantic. Auxiliary semantic chunk is optional object chunk (GBK) and the Eigen chunk (EK). The EK status is very special, it contains the statement of semantic information, determines sentence, equivalent to the grammatical predicate, mostly is the statement of the verb. Therefore, accurately judgment of the semantic feature block and its types is essential for correct understanding of statement.

2.3. Sentences Understanding Technology of HNC

HNC language understanding technology which is also called sentence category analysis can be divided into three parts: semantic block perception and sentence hypotheses, sentence test, semantic block analysis. Semantic chunk usually consists of core part and part, and EK is no exception [8].that is to say, EK is not a predicate verb, but a structural body, with composite structure. The EK core portion of the front and rear can have that part. That part is called tops, after that part is called bottoms. English EK constitution aims to facilitate computer perception to confirm EK sentence hypothesis, on the basis of translation selection strategy. Semantic chunk perception is based on the concept of dynamic (v) concept, known as the v criterion. V criterion used as guidelines for EK perception. Because the v concept will form EK, it provides information for EK .The main verb of English is v concept, the verb processing become EK perception in key. English description of EK part is located before the verb, namely top EK is widespread, and the bottoms are rare in English. Reliability is one of the most important characteristics of test. Test retest reliability can be used two times of test scores of product moment correlation coefficient formula (2) to express.

$$R_n = \frac{\sum (X - X')(Y - Y')}{\sqrt{\sum (X - X')^2} \sqrt{\sum (Y - Y')^2}} \quad (2)$$

$X' = \frac{1}{n}\sum_i^n X_i$, $Y' = \frac{1}{n}\sum_i^n Y_i$ are average number of the variable. The formula also can be expressed as formula (3).

$$R_n = \frac{\sum XY - \frac{1}{n}(\sum X)(\sum Y)}{\sqrt{\sum X^2 - nX^2} \sqrt{\sum Y^2 - nY^2}} \quad (3)$$

The range of correlation coefficient is shown in Figure 1. Word class conversion is the core part of bilingual machine translation engine based on the HNC theory. Word class conversion

is divided into three type including points to zero conversion, mandatory conversion and selective conversion. Zero transformation refers to the source language sentence to the target language sentence mapping, the sentence category is unchanged, such as the basic role of sentence; mandatory conversion refers to the sentence must be converted, it can't be transferred, such as the basic judgment sentences. The two types including mandatory conversion and selective conversion, there are deterministic and non-deterministic conversion. The former reflect the source and target sentence statement between one-to-one relationship, while the other reflect the source and target sentence statement between one-to-many relationship.

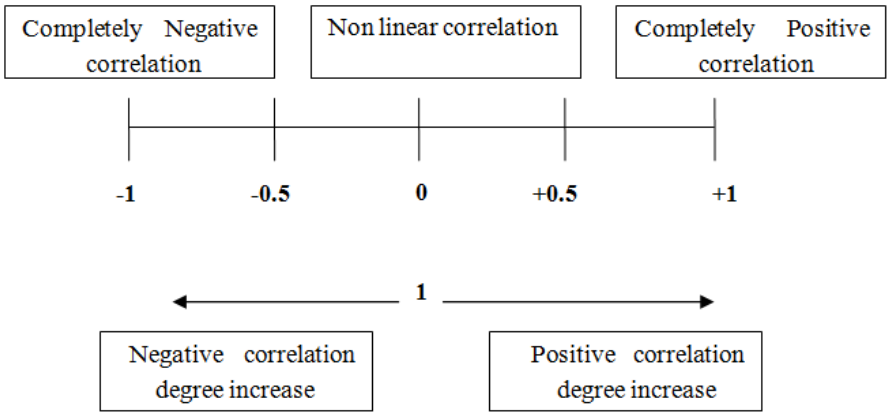

Figure 1. The range

For deterministic sentence category transformation, source language sentence correspond to a plurality of target language sentences sentence category. How to select a sentence during this limited sentence, to make the target language sentence can accurately express the source language sentence structure and semantic and conform to the habitual use of target language. The first standard is based on the corresponding to the source sentence category of the target sentence using a class degree. The use of the highest degree is preferred objects. While the using of sentence types not only rely on a high level of bilingual worker knowledge and experience, but also rely on large scale corpus and statistical techniques. Second criteria are based on the translation of target different translation objectives require different translation strategies and methods, corresponding will choose different word classes to implement the translation target.

3. Sentence Structure Analysis and English Semantic Feature Processing

3.1. Sentence Structure Analysis

The main thought of syntactic analysis is a kind of error driven method. The correction rules of error analysis are obtained through the automatic extraction and artificial participation method. After the syntactic and error analysis, correction rules are used for further processing of analysis results. English sentences have complex sentence structure, in order to analysis and conversion, structure analysis based on Extended Information-based Case Grammar is used for transformation sentence transformation. The interrogative sentence structure is the unity of the corresponding structure, and the structure transformation and the target generation when reduced to interrogative sentences. Sentence structure analysis of the complex sentence is decomposed into complex sentence structure consisted of a simple

sentence. Then the various simple sentences were processed. The analysis of simple sentence structure with the predicate verb as the centre, according to the constraints and shallow syntactic parsing results filled lattice frame, so as to get the whole sentence syntax and semantic structure. Conversion of sentences in accordance with the sentence analysis results and conversion rules to produce target language syntax and semantic structure. The logical structure of sentence structure analysis and transformation is shown in Figure2 (a) and the analysis flow chart is shown in Figure2 (b).

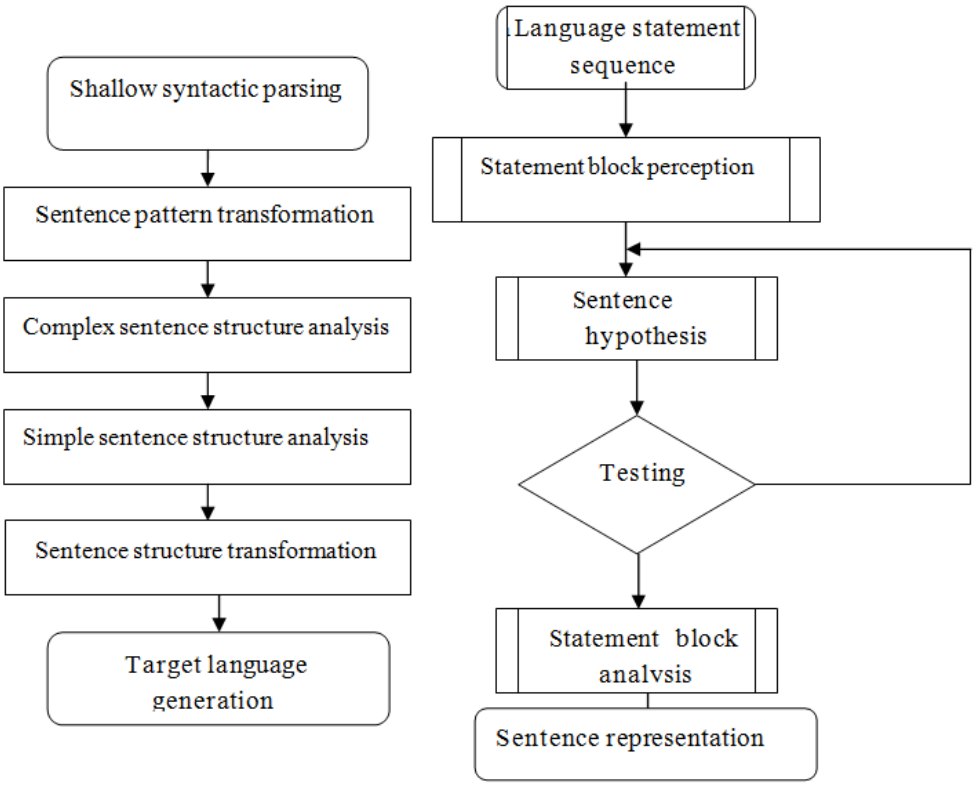

Figure 2. (a) The Logical Structure of Sentence Structure Analysis and Transformation, (b) The Analysis Flow Chart of Sentence Structure

The strategy of complex sentence processing is to break up the whole into parts. First, complex sentence are divided into a group composed of phrases and simple sentence complex sentence structure according to the sentence features, function words, punctuation tree was built. Due to the shallow parsing phase has been recognition of variety of phases, and get the syntactic structure and the translation models. Only the compound sentence, complex sentence and inserting components processing should be considered.

Simple sentence only contains a predicate or the equivalent of a verb in the verb phrase as the predicate of the sentence. Simple sentence can be described as follow:

Simple=Subject+ Predicate follow components.

Subject=Part Subject+ Subject composition+ Adverbial

Predicate=Modal words+ Auxiliary + predicate verb

Predicate follow components= Object component+ predicative constituents+ Adverbial

Subject composition = Noun phrase + Participial phase + Infinitive phrase

Adverbial= Adverb phrase + Prepositional phrases + Word Segmentation + Phrase + Infinitive phrase.

The simple sentence structure analysis use similar top-down analysis method. Firstly, find the predicate verb in a sentence, than set the predicate, and then choose the predicate as a dividing line. At last, forming the simple sentence, syntax semantic structure based on the subject and predicate of the language verb case frame analysis. English sentence only have one predicate. The predicate choose predicate verb as the center, and modals, auxiliaries and some predicate is closely linked with the modality adverbs. They used to express tense, voice and other types of syntactic features, and a sentence subject in person and number to maintain consistency, predicate analysis process is scanning the input sentence formerly backward , identified the predicate center of be verbs or true verbs , moreover there may be more than one verb or true verb used for connection. Then search of predicate boundaries forward and backward, in order to match predicate structure pattern and identifying predicate types. Finally, according to the center predicate verb and structural type, construction of predicate lattice framework and other computers of sentence constraint are carried out.

3.2. Description of EK (Eigen Chunk) Algorithm

Semantic chunks segmentation and combination is based on concept of verb criterion and language logic (LV criterion). Each word position belongs to a semantic block position information is obtained. Verb concepts belong to semantic blocks stored in the EK1 array using verb criterion. The sentence appeared in the EK semantic blocks stored in EK2 array. Because English grammar requires every sentence must have a verb, so elements in EK2 array will appear in EK1 array. To semantic chunks appear in both the EK1 array and EK2 array, priority hypothesis the EK core type for judgment. Algorithm flow diagram was shown in Figure3. It should be pointed out that, the above algorithm is put forward from the EK. In the setting of EK process also should combine the exclusion rules and queue rule. English core part of EK composite is the most complex part. The analysis of English EK composite is as follow:

Combined form: the EK consists of two or more juxtaposed with as EK status of verbs. E consist EK are located in the same layer concept node table. Words consist EK use a comma or “and” for separation. For combined EK form, computer can analysis sentence based on any type of E. For example: We sang and talked all night.

Combination type: constitute the EK words HNC symbol is not located at the same layer of concept node table. The verb will have a comma or the “and” separated marks, some explicit independent dynamic concept, each having its commonly used sentence category and paired with their respective semantic chunk. For example: I went to a supermarket, bought some drinks and then left.

Dynamic collocation refers to E is a verb (dynamic). EH is a non verb (static), or combination of the two EK. In this configuration, the sentence type is determined by the EH.

Because exists in Chinese and English source word corresponding to the dynamic and static concept ,a correlation between them and the English source sentence in static and dynamic relations among concepts of equivalent. English is paired with dynamic and static concept, while the concept in Chinese corresponding dynamic concept and static conceptual relevance. For example: America pay great attention to China-Japan relationship. The software interface of the algorithm is shown in Figure 4.

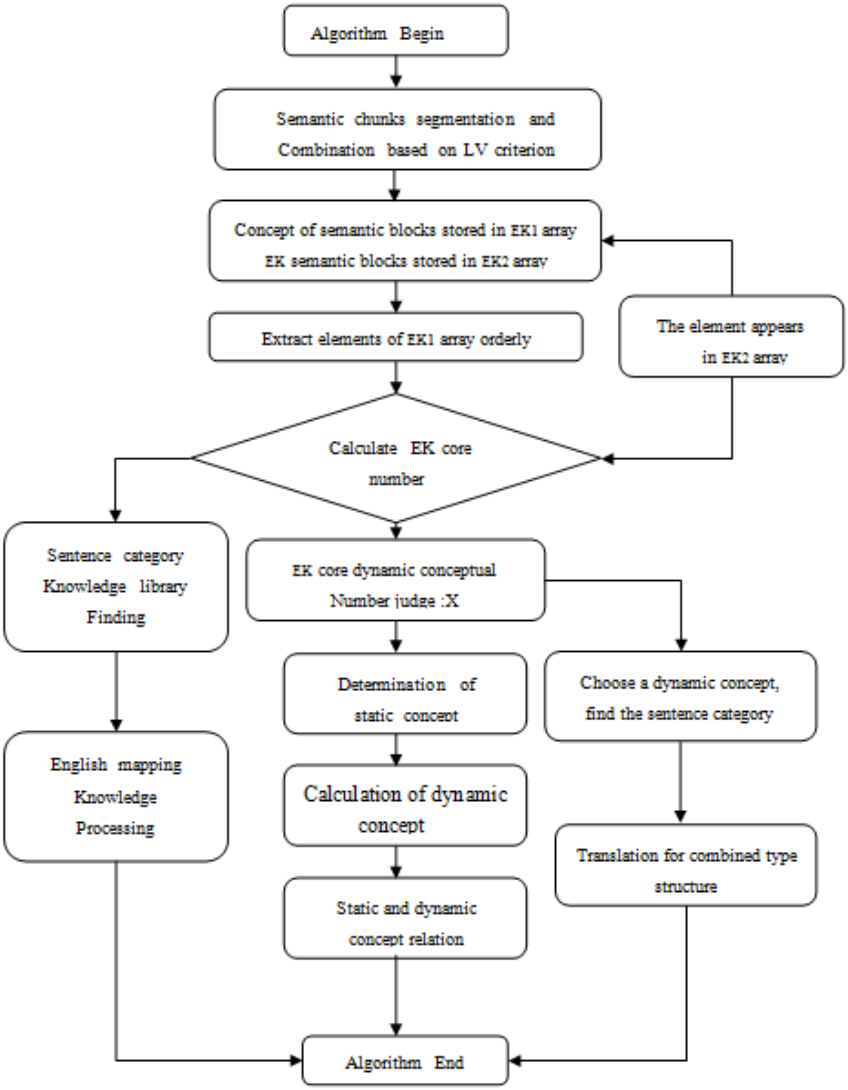

Figure 3. English Semantic Feature Processing Algorithm Flow Diagram

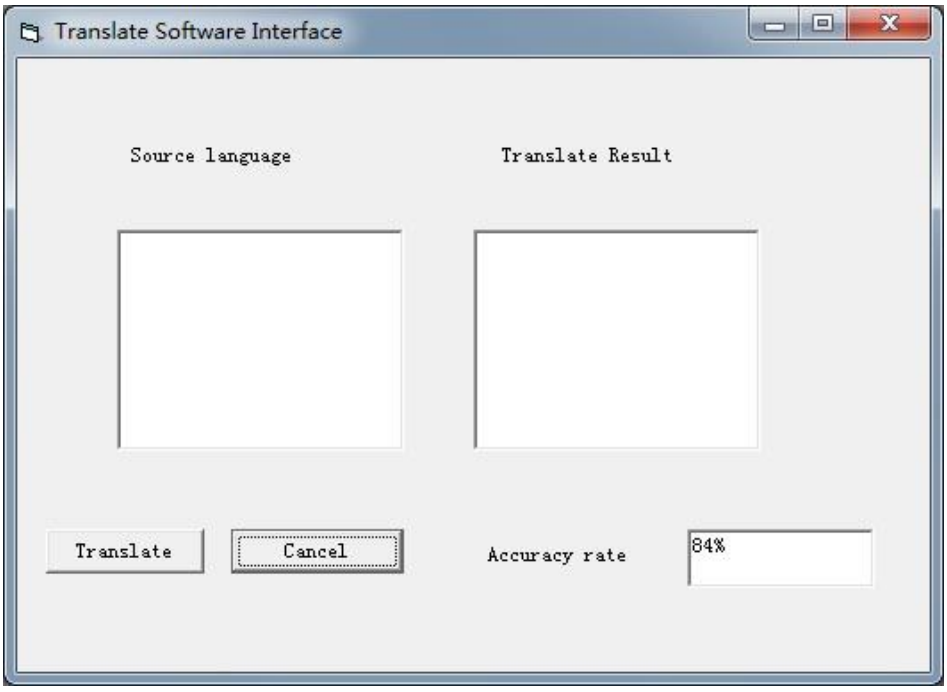

Figure 4. The Translate Software Interface of the Algorithm

3.3. Testing of EK (Eigen Chunk) Algorithm

The algorithm is tested artificially because of the limited of concept knowledge and sentence database. Sentence samples we use are form the relevant foreign website and newspaper translation, the translation is fully reflects the characteristics of the source language authoritative. We choose the online machine translation and our algorithm for experiment. The algorithm was tested and the main observation is the machine processing result is I agree to plan to withdraw from the travel. The plan as a verb in this translation, while the algorithm can judge out the plan here is the noun. When statistical the accuracy, the online machine translation is not accurate, while our algorithm is accuracy .The results of comparison between our algorithm and online machine translation are shown in table 1.

The common weakness of this algorithm and online translation machine is the English be verb processing, no expected this knowledge can be used and the accuracy is low. For this algorithm with a machine translation taste is a desired outcome. Because the machine is equipped whit a variety of translation knowledge and skills, they can only on based on the source sentence syntactic and semantic information for analysis, but this translation is acceptable.

As one of the most important aspect of HNC theory application, translation machine need to use the source language sentence category knowledge. While EK is required in order to make the correct decision of the sentence category knowledge activation. In our study, through the algorithm, and with actual comparison show that, the algorithm of the source statement were more in-depth analysis and understanding, to further improve the English-Chinese machine translation accuracy, provide technical support for machine translation.

Table 1. The Results of Comparison Between our Algorithm and Online Machine Translation

	Our algorithm	**Online machine translation**
Advantages	Identify the static and dynamic concept at low level and sentence; understand semantics accurately; process more accurately for polysemous verb by using the concept of the association.	Translate verb phrase commonly; Consist of sentence structure; Conform to Chinese language habit.
Disadvantages	Sentence structures do not conform to the Chinese usage.	Cannot accurately judge true verbs; Especially for verbs with multiple means; literal translation method in English special verb processing is poor.
Common disadvantages	Processing of "be" verb	
Characteristic	Further analysis of EK	Using statistical methods or literal translation
Accuracy rate	84%	58%

4. Conclusions

In the study, we discuss the core structure of English semantic respectively and put forward the corresponding computer algorithm. In view of the current HNC theories on English EK research, a detailed analysis of the English constitution of EK was carried out, based on the structural characteristics of the computer processing strategy. Effect chain and judge is called generalized function of sentence semantic classification. The results show this algorithm is more in-depth analysis and understanding of source sentence compared with online machine translation machines. The experimental results can provide technical support for English-Chinese machine translation. The current machine translation software usually analysis sources statement and no depth to the semantic level, which resulting in the accuracy of current machine translation. The understanding of language thought based on HNC theory combines syntactic with semantic layer face for language comprehension. Analysis of English semantic feature block structure based on the HNC theory was carried out, we proposed an English semantic feature block computer processing algorithm, the algorithm provides deeply understanding of sentence and better translation of the meaning.

References

[1] B. Wu, Y. Guo and B. Wang, "English Chinese machine Translation Rule based sentence structure Analysis and Transformation", Journal Information Engineering University, vol. 8, no. 1, (**2007**), pp. 30-33.
[2] J. Liao and Q. Zhang, "HNC Theory New Progress of Statement Format", Computer science, vol. 33, no. 5, (**2006**), pp. 173-177.

[3] KeliangZhang and Z. Huang, "HNC effect of Chinese-English Sentence Translation", Journal of Chinese Information Processing, vol. 17, no. 5, (**2003**), pp. 19-26.
[4] X. Dai, C. Yin, J. Chen and G. Zheng, "Current Situation and prospect of Research on Machine Translation", Computer Science, vol. 31, no. 11, (**2004**), pp. 176-184.
[5] Y. Meng and T. Z. X. Chen, "Based on the Evaluation of the English Syntactic Structure Disambiguation and Self-Evaluation Rule Correction", Journal of Computer and Development, vol. 39, no. 7, (**2002**), pp. 802-808.
[6] H. Yanhuang and Z. Xiongchen, "Based on Analysis of Complex Long Sentence Translation Algorithm", Journal of Chinese information processing, vol. 16, no. 3, (**2002**), pp. 1-2.
[7] C. Liu and C. Wu, "Sentence Decomplexification using Holistic Aspect-based Clause Direction for Long sentence Understanding",7th International Symposium on Chinese Spoken Language Processing (ISCSLP), (**2010**), pp. 265-270.
[8] X. Sun, F. Ren and D. Huang, "Extended Super Function based Chinese Japanese Machine Translation", International Conference on Natural Language Processing and Knowledge Engineering, (**2009**), pp. 1-8.
[9] X. Dong, H. Xue, and Y. Yang, "Factor-based Uyghur-Chinese statistical Machine Translation" IJACT: International Journal of Advancements in computing Technology, vol. 4, no. 2, (**2012**), pp. 275-283.
[10] W. Li and A. Pei, "The study on English-Chinese Machine Translation based on Date-Oriented Parsing Theory", AISS: Advance in Information Sciences and Service Sciences, vol. 4, no. 1, (**2012**), pp. 69-76,.
[11] Q. Wang, H. Ma, Y. Chi, Y. Li, L. Dong and D. Wang "Chinese Word Knowledge Improvement based on HNC", Journal of Chinese Information Processing ,vol. 26, no. 2, (**2012**), pp. 35-39.
[12] P. Benson, "Teaching and Researching Autonomy in Language Learning", Person Education Press, China, (**2001**).
[13] S. G. Paris and L. R. Ayres, "Becoming Reflective Students and Teacher", American Psychological Association, USA, (**1994**).
[14] T. Anderson, and F. Elloum, "Theory and Practice of Online Learning", Athabasca University Press, Canada, (**2005**).
[15] Z. Yangen and S. Qingsong, "The study on the multi-media English teaching of colleges", Foreign Language Teaching and Study, vol. 9, no. 2, (**2003**), pp. 43-49.
[16] D. Chun, "Using computer networking to facilitate the acquisition of interactive competence", System, (**1994**)s.
[17] J. S. Lamancusa, J. E. Jorgensen, J. L. Zayas-Castro and J. Ratner, "The Learning Factory- A New Approach to Integrating Design and Manufacturing into Engineering Curricula", 1995 ASEE Conference Proceedings, (**1995**), June 25 28, Anaheim, CA, pp. 2262-2269.
[18] C. L. Hidalgo and J. R. Williams, "WEB-ducation: Extending a Teacher's Communication and Mediation Capabilities through the Internet", Engineering Education Innovators Conference, sponsored by NSF, April 7-8, 1996, Washington D.C., URL: http://monett.mit.edu/nsf/trpdoc.nsf.
[19] S. R. Lerman and J. N. Lapierre, "A Multimedia Model on Statistics in Manufacturing Quality Control", Engineering Education Innovators Conference, sponsored by NSF, (**1996**), April 7-8, http://www-ceci.mit.edu/projects/manufacturing/CEPsummary.html
[20] P. L. Jackson, "OPTLINE: Manufacturing Process Line Simulation", Engineering Education Innovators Conference, sponsored by NSF, (**1996**), April 7-8, Washington D.C., URL: http://www.orie.cornell.edu/~jackson/optline.html.
[21] J. E. Wood, H. Hahn, P. Kunsberg, H. Ravinder and J. N. Beer, "The UNM Manufacturing Engineering Program: Manufacturing Enterprise Simulator", Engineering Education Innovators Conference, sponsored by NSF, (**1996**), April 7-8, Washington D.C., URL: http://www-mep.unm.edu/Paper97_TRP1221.htm
[22] J. S. Lamancusa, M. Torres, V. Kumar and J. Jorgensen, "Learning Engineering by Product Dissection", (**1996**) ASEE Conference Proceedings, June 23-26, 1996, Washington, DC.
[23] L. Morrel, J. Zayas, J. S. Lamancusa and J. E. Jorgensen, "Making a Partnership Work: Outcomes Assessment of a Multi-Institutional, Multi-Task Project", Engineering Education Innovators Conference, sponsored by NSF, April 7-8, 1996, Washington D.C., URL: http://mayaweb.upr.clu.edu/Paper/techno.html
[24] Authorware from Macromedia, Inc., 600 Townsend St., San Francisco, Ca. 94103.

Author

Bangqing Pei, he received his B.A (1998) and M.Ed in (2005) from University. Now he is full lecturer at School of Foreign Languages, Mianyang Normal University, Mianyang, China. Since 1998 he has been an English teacher in Mianyang Normal University. His current research interests include different aspects of English language teaching.

