We are IntechOpen, the world's leading publisher of Open Access books
Built by scientists, for scientists

6,900
Open access books available

185,000
International authors and editors

200M
Downloads

154
Countries delivered to

Our authors are among the

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Cognitive Learning for Sentence Understanding

Yi Guo, Zhiqing Shao
Department of Computer Science and Engineering
East China University of Science and Technology
China

1. Introduction

In the research field of natural language understanding, sentence stands a very prominent position in text processing. The process of sentence understanding involves computing the meaning of a sentence based on analysis of meanings of its individual words. Research procedures in sentence understanding examine the representations and processes that connect the identification of individual words in text reading (Culter, 1995; Balota, 1994) with mapping sentence meanings to relevant mental models (Johnson-Laird, 1983) or discourse representations (Kintsch, 1988; van Eijck & Kamp, 1997).

The task of sentence understanding includes two stages, sentence parsing and semantic processing. Sentence parsing resides in the fundamental level, while semantic understanding involves lexial and higher discourse analysis. Sentence understanding has compact connections with human cognition, thus this chaper will introduce how cognitive models are integrated, with machine learning algorithms (or models), into the procedures of sentence parsing and semantic processing.

2. Statistical Learning Review

Over the past decade, statistical learning, a means to discover hidden structures or patterns by analyzing statistical properties of the input, has emerged a general candidate mechanism by which a wide range of linguistic experience can be acquired (Saffran, 2003).

Statistical learning, as a type of implicit learning, has been demonstrated across a variety of natural and artificial language learning situations, including learning of information that is potentially highly relevant to sentence comprehension processes, such as using function words to delineate phrases (Green, 1979), integrating prosodic and morphological cues in the learning of phrase structure (Morgan et al., 1987), parsing each natural language sentence (Charniak, 1997) to a hierarchical structure which presents how words hookup together to form constituents, discovering phonological and distributional cues to lexical categories (Monaghan et al., 2005), locating syntactic phrase boundaries (Saffran, 2001; 2003), and detecting long-distance relationships between words (Gómez, 2002; Onnis et al., 2003).

Misyak & Christiansen (2007) revealed that statistical learning ability was a stronger predictor of relative clause comprehension than the reading span measure, and suggested that statistical learning may play a strong role in the accumulation of linguistic experience relevant for sentence processing.

Moreover, within natural language comprehension and production studies, there is clear evidence that prior experience of a given syntactic structure affects (1) comprehension of similar structures and (2) the probability that a speaker will utter a sentence with the same or similar structure, even when there is no meaning overlap between sentences (Ferreira & Bock, 2006).

Syntactic priming has been described as stemming from statistical learning at the syntactic level (Bock & Griffin, 2000; Chang et al., 2006) or at the syntactic–semantic interface (Chang et al., 2003), which can be viewed as examples of statistical learning of information relevant to sentence processing.

Above research works have testified the significance of statistical learning for natural language processing, including sentence comprehension, and also explicitly pointed out the performance bottlenecks (Monaghan et al., 2005; Dell & Bock, 2006; Misyak & Christiansen, 2007) of statistical processing technologies. Since human, rather than the computer software and hardware, is the core subject to process and understand natural language, it is essential to survey pivotal research works regarding human cognition.

3. Cognitive Concepts Highlight

Sentences convey not only lexico-semantic information for each word, but sentence meaning based on syntactic structures (Townsend & Bever, 2001; Friederici, 2002), which has elucidated the importance of syntactic structures for sentences.

Recursion is a unique human component of the faculty of language (Hauser et al., 2002), which is also known as the property of discrete infinity, the ability to generate an infinite range of discrete expressions from a finite set of elements. Sentences are indeed such infinite expressions generated from a limited set of words, signs, or letters; and syntactic mechanisms (Chomsky, 2000) have been applied to instantiate this property.

Thus, the processing of syntactic structures plays a critical role in the selective integration of lexico-semantic information into sentence meaning. Syntactic analyses are performed in the service of semantics, and sentence meaning is derived from syntactic analyses of the sentence structures.

As mentioned before, the procedure of sentence understanding includes sentence parsing at a fundamental level, and semantic understanding at lexial and higher discourse analysis. This section will highlight several cognitive concepts regarding sentence parsing and emantic understanding.

3.1 Syntax-First and Interactive Models

How human beings parse sentences, especially for syntactically ambiguous sentences, has been a long-history cognitive research topic attracting research efforts for decades in the field of cognitive psychology. In cognitive psychology, behavioristic experiments have been popularly implemented to explore the sentence-analyzing mechanism, which is also called "parser", especially in the case that human beings cannot automatically constitute the meaning of a sentence.

With respect to syntactic and semantic processing in sentence comprehension, two main classes of cognitive models have been proposed to account for the behavioral data: Syntax-First and Interactive models.

Syntax-First models (Fodor, 1983; Frazier & Fodor, 1978; Kako & Wagner, 2001) claims that, (1) syntax plays the main part whereas semantics is only a supporting role, (2) the parser initially builds a syntactic structure based on word category information, which is independent from lexical or semantic information, and (3) thematic role assignment takes place during a second stage. If the initial syntactic structure cannot be mapped onto the thematic structure, the final stage will require a re-analysis.

Interactive models (Bates & Mac-Whinney, 1987; MacDonald et al., 1994; Marslen-Wislon & Tyler, 1980; Taraban & McClelland, 1988) state that syntactic and semantic processes actually interact with each other at an early stage, and both syntax and semantics work together to determine the meaning of a sentence. Despite the agreement that syntactic and semantic information has to be integrated within a short period of time, the two model classes differ in their views on the temporal structure of the integration processes.

Syntax and semantics are two indispensable properties of sentences. The eye-tracking studies (Tanenhaus & Trueswell, 1995) have supported the conclusion that syntax and semantics interact during parsing, which denotes that meaning affects early processing. These behavioristic experiments have convinced that the interactionist approach (Trueswell et al., 1994) is rational and effective to simulate human parsing and semantic understanding mechanism.

3.2 The Garden Path Model and Alternatives

Theories of sentence processing have illustrated various perspectives on when comprehenders initiate semantic interpretation of an incoming word. One of the most prominent and influential models of sentence processing is the garden path model (Ferreira & Clifton, 1986; Frazier & Rayner, 1982; Rayner et al., 1983), which states that semantic interpretation generally followes the construction of a syntactic analysis.

The syntactic analysis applies appropriate syntactic parsing strategy together with information of major syntactic category (e.g. noun, verb, adjective, etc.) of incoming words. Semantic interpretation can proceed after he construction of a syntactic analysis. The strict temporal ordering of syntactic analysis and semantic interpretation produce the fact that semantic information cannot influence the construction of a syntactic analysis. The effects of semantic information observed during the resolution of syntactic ambiguity have been interpreted as reflecting processes occurring after an initial syntactic analysis (Ferreira & Clifton, 1986; Kennison, 2001; Speer & Clifton, 1998).

From the perspective of the garden path theory, the results of the present research can be viewed as supporting the claim that certain aspects of high-level integrative semantic processing for an incoming word occur only after the comprehender determines the word's syntactic analysis.

The most prominent alternatives to the garden path model include interactive and constraint-based approaches to sentence processing. These approaches stated that language comprehension can be achieved through highly interactive and parallel processing (MacDonald et al., 1994; Sedivy et al., 1999; Tanenhaus et al., 1995; Taraban & McClelland, 1988; Trueswell & Tanenhaus, 1994; Trueswell et al., 1993). Word-specific (lexical) information will produce candidate syntactic frames, which are activated in parallel.

Although semantic interpretations are constructed upon syntactic frames (MacDonald et al., 1994), semantic information can influence the activation of syntactic frames. As a consequence, syntactic and semantic analysis may influence each other.

3.3 The Brain-Based Model

In the brain-based model (Friederici, 2002), language comprehension is divided into three functionally and temporally separable processing steps: (1) initial local structure building in the first phase; (2) lexical-semantic and thematic processes in the second phase; and (3) syntactic integration and revision in the third phase. For an integrative view of language processing, recent brain image research (Friederici & Kotz, 2003) provides support evidence that syntax-first aspects take place in an early time window and the interactive aspects happen in a later time window.

3.4 Working Memory and Semantic Memory

An early study (Fodor, 1983) of sentence understanding hypothesized a cognitive architecture focusing on a component building grammatical structures of sentence processing. Later research works (Caplan & Waters, 1999; Gibson, 1998; Just & Carpenter, 1992; Zurif et al., 1995) involve various executive resources facilitating sentence processing, such as working memory (WM), which contains specific sentence features and acts as temporary storage for phrasal information manipulation during the processing of long-distance syntactic dependencies in a sentence. During the course of sentence processing, working memory may help maintain, in a linear or non-linear manner, crucial components of a sentence in an active state until the correct grammatical relationships are established (Lewis et al., 2006).

Tulving (1972) first introduced semantic memory (SM), which refers to the general knowledge of concepts and facts, including word meaning, and involves encoding and retrieval of information in multiple domains (Hart et al., 2007). The essence of semantic memory is that contents are not statically bound to any particular instance of experience as in episodic memory. Instead, semantic memory stores is the gist of experience, an abstract structure applicable to a wide range of experiential objects, and delineates categorical and functional relationships between such objects.

4. Simple Recurrent Networks (SRNs)

Hadley (1994) proposed that systematic behavior is a matter of learning and generalization; thus, a neural network trained on a limited number of sentences should to be able to process all possible sentences in a generalize manner. Moreover, since people learn systematic language behavior from exposure to only a small fraction of possible sentences, a neural network should similarly be able to learn from a relatively small proportion of possible sentences, if it is to be considered cognitively plausible.

Simple Recurrent Networks (SRNs) (Elman, 1991) has been widely applied in basic connectionist approaches (parallel distributed processing) for language learning. SRN has been implemented to employ the functions of working memory (MacDonald et al., 2001; MacDonald & Christiansen, 2002).

The SRN architecture (as illustrated in Fig. 1.) includes the activations from the recurrent layer (RL, the hidden layer) as the context layer (CL) in the input layer (IL), aiming at processing inputs that consist of sequences of patterns of variable length. This architecture allows the network to include information connected with all the previous steps in a sequence in its processing of the current stage. The architecture will remember what has gone before, forgetting gradually as it progresses through the sequence.
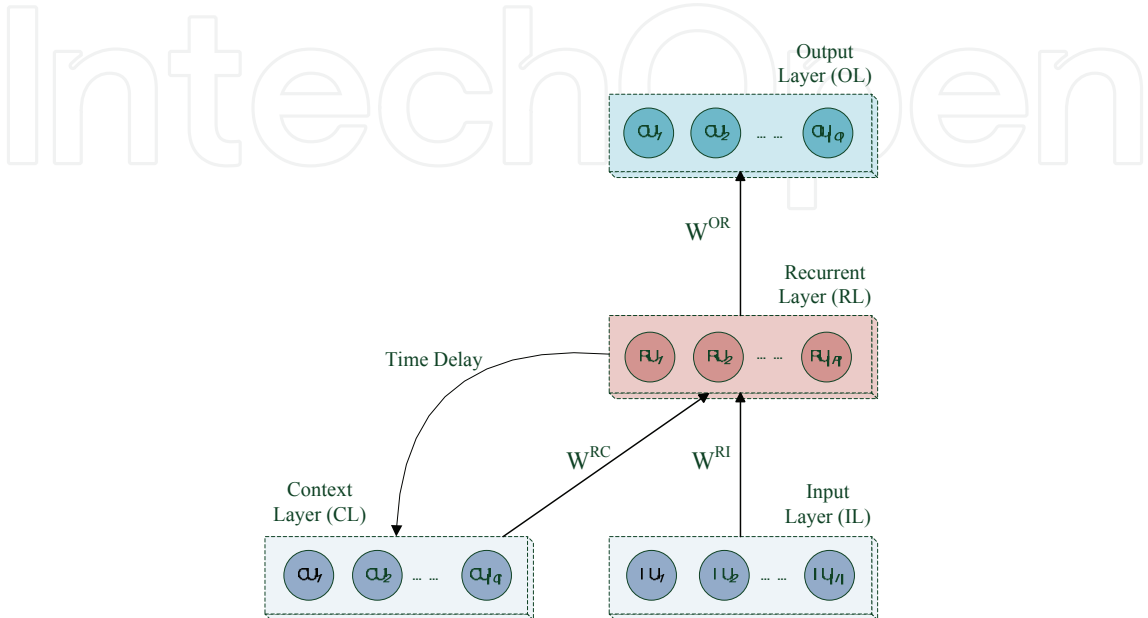
Fig.1. Architecture of Simple Recurrent Networks

Symbols	Definition
IU	A unit of input layer
RU	A unit of recurrent layer
CU	A unit of context layer
OU	A unit of output layer
$|I|$	The number of units in IL
$|R|$	The number of units in RL
$|C|$	The number of units in CL
$|O|$	The number of units in OL
W^{RI}	The weight vector from IL to RL
W^{RC}	The weight vector from CL to RL
W^{OR}	The weight vector from RL to OL

Table 1. Definition of SRN Symbols

Symbols in Fig. 1. are defined in table 1: the first order weight matrices W^{RI} and W^{OR} fully connect the units of the input layer (IL) , the recurrent layer (RL) and the output layer (OL) respectively, as in the feed forward multilayer perceptron (MLP). The current activities of recurrent units $RU^{(t)}$ are fed back through time delay connections to the context layer, which is presented as $CU^{(t+1)} = RU^{(t)}$.

Therefore, each unit in recurrent layer is fed by activities of all recurrent units from previous time step through recurrent weight matrix W^{RC}. The context layer, which is composed with

activities of recurrent units from previous time step, can be viewed as an extension of input layer to the recurrent layer. Above working procedure represents the memory of the network via holding contextual information from previous time steps.

The weight matrices W^{RI}, W^{RC} and W^{OR} are presented as equations (1) to (3)

$$W^{RI} = [\,(w_1^{RI})^T, (w_2^{RI})^T, ..., (w_{|R|}^{RI})^T\,] = \begin{bmatrix} w_{11}^{ri} & w_{12}^{ri} & \cdots & w_{1,|R|}^{ri} \\ w_{21}^{ri} & w_{22}^{ri} & \cdots & w_{1,|R|}^{ri} \\ \vdots & \vdots & \ddots & \vdots \\ w_{|I|,1}^{ri} & w_{|I|,2}^{ri} & \cdots & w_{|I|,|R|}^{ri} \end{bmatrix} \tag{1}$$

$$W^{RC} = [\,(w_1^{RC})^T, (w_2^{RC})^T, ..., (w_{|R|}^{RC})^T\,] = \begin{bmatrix} w_{11}^{rc} & w_{12}^{rc} & \cdots & w_{1,|R|}^{rc} \\ w_{21}^{rc} & w_{22}^{rc} & \cdots & w_{1,|R|}^{rc} \\ \vdots & \vdots & \ddots & \vdots \\ w_{|C|,1}^{rc} & w_{|C|,2}^{rc} & \cdots & w_{|C|,|R|}^{rc} \end{bmatrix} \tag{2}$$

$$W^{OR} = [\,(w_1^{OR})^T, (w_2^{OR})^T, ..., (w_{|O|}^{OR})^T\,] = \begin{bmatrix} w_{11}^{or} & w_{12}^{or} & \cdots & w_{1,|O|}^{or} \\ w_{21}^{or} & w_{22}^{or} & \cdots & w_{1,|O|}^{or} \\ \vdots & \vdots & \ddots & \vdots \\ w_{|R|,1}^{or} & w_{|R|,2}^{or} & \cdots & w_{|R|,|O|}^{or} \end{bmatrix} \tag{3}$$

In above formulations, where $(w_k^{RI})^T$ is the transpose of w_k^{RI} for the instance of W^{RI}, where w_k^{RI} is a row vector, $(w_k^{RI})^T$ is the column vector of the same elements. The vector $w_k^{RI} = (w_{1k}^{ri}, w_{2k}^{ri}, ... w_{|I|,k}^{ri})$ represents the weights from all the input layer units to the recurrent (hidden) layer unit RU_k. The same conclusion applies with W^{RC} and W^{OR}

Given an input pattern in time t, $\mathbf{IU}^{(t)} = (IU_1^{(t)}, IU_2^{(t)}, ..., IU_{|I|}^{(t)})$, and recurrent activities $\mathbf{RU}^{(t)} = (RU_1^{(t)}, RU_2^{(t)}, ..., RU_{|R|}^{(t)})$), for the ith recurrent unit, the net input $RU_i'^{(t)}$ and output activity $RU_i^{(t)}$ are calculated as equations (4) and (5).

$$RU_i'^{(t)} = \mathbf{IU}^{(t)} \cdot (w_i^{RI})^T + \mathbf{RU}^{(t-1)} \cdot (w_i^{RC})^T = \sum_{j=1}^{|I|} IU_j^{(t)} w_{ji}^{ri} + \sum_{j=1}^{|R|} RU_j^{(t-1)} w_{ji}^{rc} \tag{4}$$

$$RU_i^{(t)} = f(RU_i'^{(t)}) \tag{5}$$

For the kth output unit, its net input $OU_k'^{(t)}$ and output activity $OU_k^{(t)}$ are calculated as equations (6) and (7).

$$OU'^{(t)}_k = \mathbf{RU}^{(t)} \cdot (w^{OR}_k)^T = \sum_{j=1}^{|R|} RU^{(t)}_j w^{or}_{jk} \tag{6}$$

$$OU^{(t)}_k = f(OU'^{(t)}_k) \tag{7}$$

Here, the activation function f applies the logistic sigmoid function (Eq. 8).

$$f(x) = \frac{1}{1 + e^{-x}} = \frac{e^x}{1 + e^x} \tag{8}$$

5. Cognitive Learning with Machine

Most current cognitive models of language processing agree that sentence comprehension involves different types of constraints (Jackendoff, 2002) in which syntactic and semantic (conceptual) information deserve the most salient consideration.

From one point of view, separable, independent but partly sequential processes construct distinct syntactic and semantic representations of a sentence (Berwick & Weinberg, 1984; Ferreira & Clifton, 1986). The opposed view is that syntactic and semantic constraints directly and simultaneously interact with each other at the message-level representation of the input (Johnson-Laird, 1983; Marslen-Wilson & Tyler, 1987; McClelland et al., 1989).

There also exist other proposals in between above fully independent and fully interactive models. Frazier (1987) suggests that syntactic analysis is autonomous and independent from semantic variables in initial stage(s), but is affected by semantic variables at later stage(s); in the contrary, syntactic analysis can influence semantic integration from the very beginning of processing. Meanwhile, Trueswell et al., (1994) claims that semantic information affects and leads syntactic analysis of the utterance in an immediate and direct mannner.

Abve and several other diverging proposals can be testified by using event-related brain potentials (ERPs), measurements of brain activity, which are elicited during the process of sentence comprehension. Different reliable ERP components have been employed to prove the distinction between the processing of syntactic and semantic information during sentence understanding. The extent and type of interaction of ERPs are taken as evidence for the interplay occurring between syntactic and semantic analyses during sentence comprehension.

This section will focus on how syntactic parsing and semantic processing are implemented with sentence processing machinery.

5.1 Sentence Parsing

The processing of syntactic structures plays a critical role in the selective integration of lexico-semantic information into sentence meaning. Syntactic analyses are performed in the service of semantics, and sentence meaning is derived from syntactic analyses of the sentence structures.

As discussed in section 3.1, the behavior experiments proved that semantics and syntax work together in sentence parsing to clarify the meaning. As a conclusion, semantics should be assigned an equal prominent role as syntax to improve parsing results. Thus, how to incorporate semantics with syntax simultaneously is the dominant challenge in sentence parsing.

In recent a few years, the research works of natural language processing (NLP) have strived toward the elaboration of huge linguistic dictionaries and ontologies (Knight et al., 1995; Miller et al., 1990; Sugumaran & Storey, 2002), even including relations between concepts and common sense. The exploitation and implementation of such dictionaries and ontologies has fulfilled some *understanding* requirements.

Kapetanios et al. (2005) proposed to implement the process of parsing natural language queries with an ontology, which preserved extensional semantics, such as domain terms, operators and operations. Since the context of terms circumscribed by the real-world semantics can be expressed by the ontology, it also will alleviate the semantic parsing. Context of terms is defined by the interrelationships expressed with an ontology as well as by the intentional meaning expressed with annotations.

Considering the impacts of linguistic dictionaries and ontologies in NLP, our solution for interactionist parsing, CIParser, takes WordNet (Miller et al., 1990) as the linguistic dictionary, and designs a corresponding ontology, *WNOnto* (as defined in Guo & Shao (2008)), referring to a W3C working draft (van Assem et al., 2006). Since nouns and verbs are more dominant in parsing sentences into phrases, they are the word types deliberately chosen for semantical analysis with WordNet. Therefore, the design of WNOnto grounds on nouns and verbs, which also benefits time efficiency in machine learning and parsing.
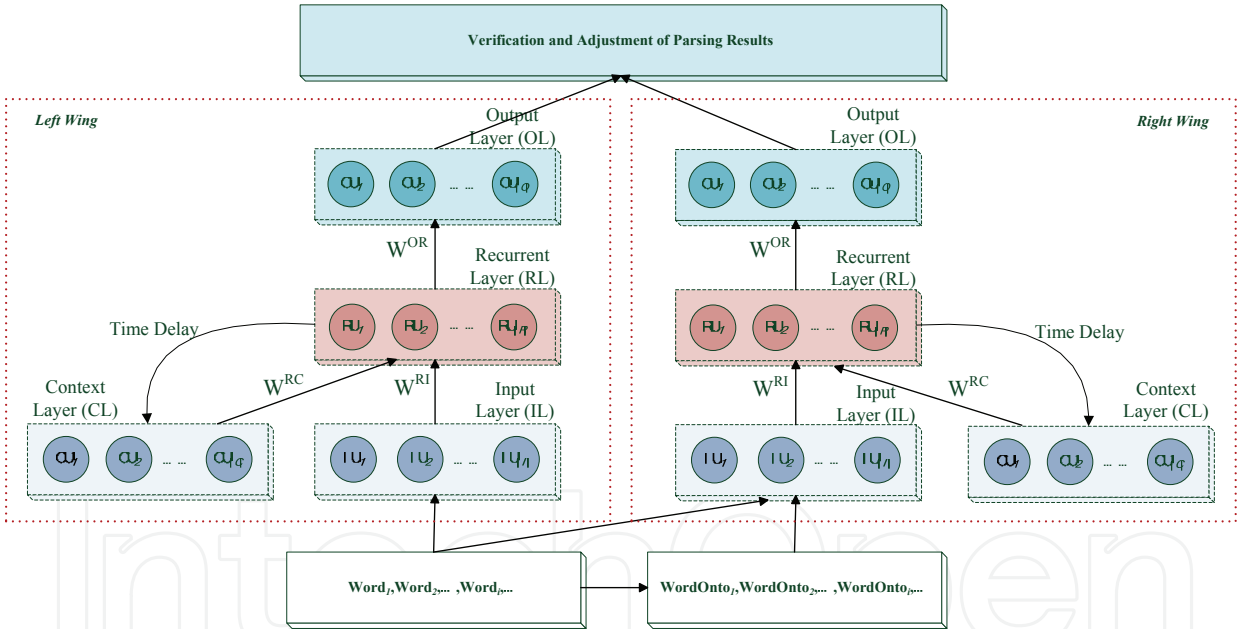

Fig. 2. Architecture of CIParser

Based on the architecture of SRN (figure 1), our CIParser is designed as illustrated in Fig. 2. The left wing is a classical SRN as described in section 4: all the input units in IL are single words from original sentences; the activations from RL of the previous time step produce the CL for the current stage; the units of IL and CL respectively multiplying matrices of W^{RI} and W^{RC} compose the input of RL; the activations of RL multiplying W^{OR} produce the input units of OL in current stage.

All the grammatical information is implicitly preserved in its pattern of link weights. Moreover, there are fewer independence assumptions. The SRN itself decides what to pay

attention to and what to ignore. Statistical issues, such as combining multiple estimators or smoothing for sparse data, are handled in the training procedure. "One-size-fits-all" is a common feature of machine learning techniques.

The right wing is structurally identical as the left wing, except that the input units in IL include not only single words from sentences but also individual ontologies, *WordOnto*s, produced according to WNOnto with querying results from WordNet. In another word, each input unit of IL is composed with (1) a single word and (2) a corresponding ontology (only for a noun or verb). Here, any noun or verb has been appended with its semantical information from WordNet in the ontology manner.

The syntactic structure of a natural language sentence is a hierarchical structure, which represents how the words connect together to form constituents, such as phrases and even clauses. This structure is normally specified with a constituent-tree, in which the constituents are nodes or leaves and the hierarchical structure is denoted with parent-child relationships.

In the final processing phase, "Verification and Adjustment of Parsing Results", the parsing results of left and right wings are verified against each other in case that either wing takes too long time to deliver a parsing result. In the case of both wings producing parsing results, we have followed a selection rule that the tree containing more constituents wins, which has been strictly followed in later experiments. The application of phrases to identify structural constituents in our CIParser also offers the competence to generalize machine learned information across structural constituents.

As we know that (1) people has language processing constraints in constructions, such as center embedding (Chomsky, 1959), and (2) people can only activate a limited number of information units in memory at any one time (Miller, 1956), we introduced working memory (Baddeley & Susan, 2006) into our CIParser. Baddeley et al. (2006) defined working memory as a limited capacity system for temporary storage and manipulation of information for complex tasks such as comprehension, learning and reasoning. In this paper, we add the storage task of working memory to our CIParser to simulate human processing features.

The nature of SRN decides that each new input, a word or/and its ontology, of the network, will also be input of the network in a new state, which indicates that information is computed through all of these states in every subsequent time period. However, the constraints on the depth of center embedding (Chomsky, 1959) implies that a limited number of these states will be referred to by following parts of the constituent-tree in any given time period.

In CIParser, we construct a queue with limited length to simulate the active units in human memory. When the SRNs arrive at a new state, this state will be queued from head to tail. When a new state comes to the queue fully filled with previous network states, the oldest state leaves the queue at tail and the new one enters the head. This queue mechanism presents that, when the number of states exceeds the queue length, the oldest state will be forgotten. This mechanism also helps the CIParser to focus on active states and to achieve precise computing results efficiently.

Guo & Shao (2008) has designed and constructed experiments for training and examining CIParser in sentence parsing. The experiments demonstrate that the SRN-based CIParser may be used for connectionist language learning with structured output representations.

The performance of CIParser is evaluated in terms of traditional measures, *Precision* and *Recall* of constituents with the famous SUSANNE Corpus.

The experimental results demonstrate that the CIParser has comparability with the state-of-the-art parsing techniques based on statistical language learning. Guo & Shao (2008) also pointed out that (1) thinking of the parsing efficiency, only the semantic information of nouns and verbs are considered in current stage; (2) involving other word types (e.g. adverb and adjective) will be future research efforts.

5.2 Semantic Processing

As we know, several knowledge repositories, e.g. WordNet (Miller et al., 1990) and Cyc (Lenat, 2006), have been developed to support programs (or agents) to increase the intelligence of specified tasks. Meanwhile, other existing repositories are domain dependent and only represent information about certain aspects of the domains.

WordNet, as a linguistic repository, does not have the capability to capture the semantic relationships or integrity constraints between concepts. As linguistic repositories lack semantic knowledge, query expansion cannot deal with several issues: (1) knowledge related to the domain of the query, (2) common sense inferences, or (3) the semantic relationships in which the concepts of the query can participate.

The Cyc ontology is a semantic repository developed to capture and represent common sense, but can not represent linguistic relationships of the concepts (e.g. whether two concepts are synonyms). Semantic repositories need linguistic knowledge to identify relevant concepts from the repository that represent a given term used in the query. Thus, a semantic repository, as Cyc, can be extended with linguistic information from the WordNet lexicon, and factual information from the World Wide Web.

In section 5.1, we have illustrated a model for sentence parsing, and we will construct another model (as Fig. 3.) for semantic processing in this section. In order to implement semantic processing in sentence understanding, we have to consider semantic repositories to represent semantic information; the integration of linguistic and semantic information could be useful to increase the contexts where knowledge in these repositories can be used successfully.

In step one, each original sentence will be first processed by CIParser to obtain a corresponding syntactic structure, e.g. a constituent tree. In step two, as the sentence is processed word by word, open and closed class words are segregated into distinct processing streams. The Grammatical Relations Mapping module will integrate constituent information for each word or phrase with strict mapping operations. In step three, the Linguistic Relations Construction module constructs linguistic relationships (e.g. synonyms, antonyms, hypernyms, and hyponyms) of the concepts in a sentence with referent provided by WordNet. In step four, the Semantic Relations Construction module captures the semantic relationships or integrity constraints between concepts, so as to successfully deal with domain knowledge and common sense inferences. Finally, in step five, all the structured data (instances of ontolgy in XML format) from previous processing steps are used to fulfill the appropriate components of the meaning structure, the Sentence-Meaning Construction Index (SMCI). Obviously, SMCI contains four types, lexical, syntactic, grammatical and semantic (or conceptual) of information.
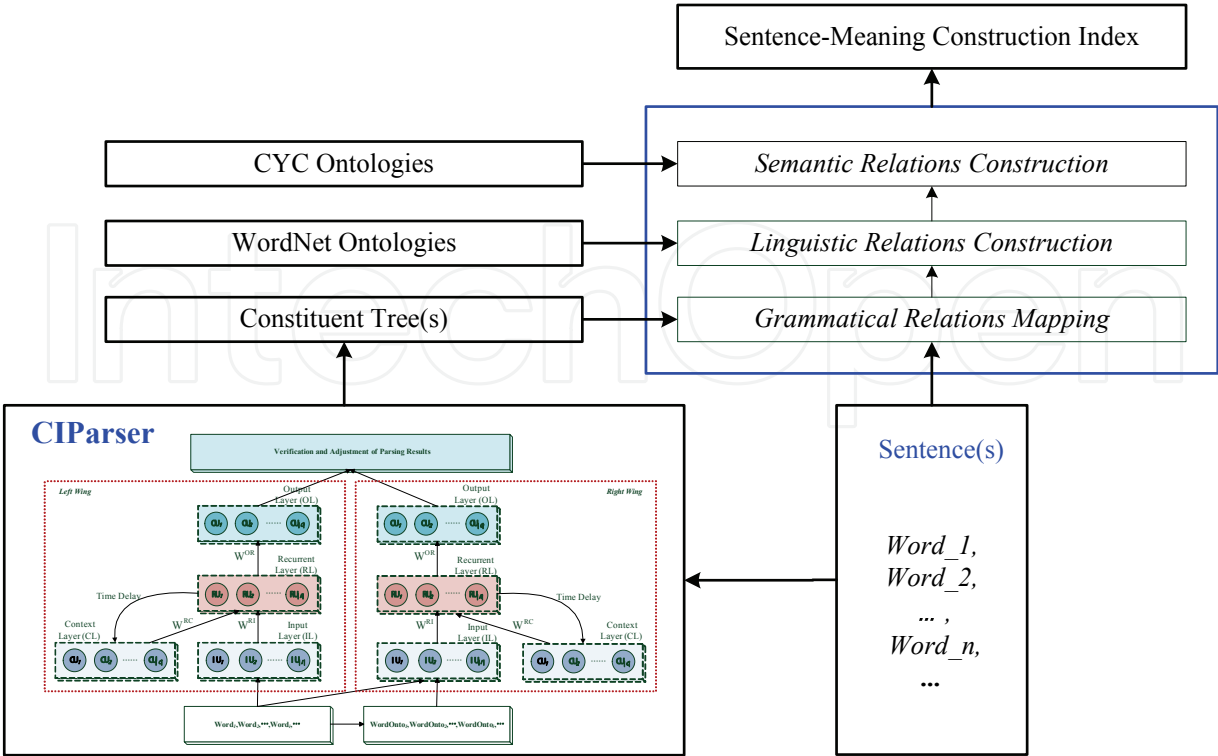
Fig. 3. A Computing Model of Semantic Processing for Sentence Understanding

The above model is able to store and retrieve different sentence-meaning construction appropriate for different sentences. The requirement is that each individual sentence should yield a unique construction index. The construction indices are used in a working memory or an associative memory to store and retrieve the correct sentence-meaning construction index.

6. Conclusion

This chaper starts with a review of classical and traditional statistical learning approaches. As sentence understanding has latent compact connections with human cognition, this chaper also highlights relevant cognitive concepts or models in sentence understanding domain. Afterwards, this chapter described the completion of sentence understanding task from two aspects, sentence parsing and semantic processing, and how cognitive models are integrated, with machine learning algorithms (or models), into the procedures of sentence parsing and semantic processing.

The CIParser has been evaluated and proven comparablr with the state-of-the-art parsing techniques based on statistical language learning. Another computing model of Semantic Processing for Sentence Understanding (Fig. 3.) also has been constructed to deliver Sentence-Meaning Construction Index (SMCI) for each sentence. With SMCI, a sentence can be understood in four dimensions, which are lexical, syntactic, grammatical and semantic (or conceptual) dimensions.

Cognitive learning with machines for sentence understanding has just started with minor productions, in which our works took SRNs as an initial model of artificial neural networks.

In an artificial language learning task (next-word prediction), van der Velde et al. (2004) evaluated a simple recurrent network (SRN) and claimed that the SRN failed to process novel sentences appropriately, for example, by correctly distinguishing between nouns and verbs. However, Frank (2006) extended above simulations and showed that, although limitations had arisen from overfitting in large networks (van der Velde et al., 2004), an identical SRN still can display some generalization performance in the condition that the lexicon size was increased properly. Moreover, Frank (2006) demonstrated that generalization could be further improved by employing the echo-state network (ESN) (Jaeger, 2003), an alternative network that requires less training (due to fixed input and recurrent weights) and is less prone to overfitting.

Recurrent Self-Organizing Networks (RSON) (Farkaš & Crocker, 2006), coupled with two types of a single-layer prediction module, had demonstrated salient benefit in learning temporal context representations. In the task of next-word prediction, RSON achieved the best performance, which turned out to be more robust and faster to train than SRN and higher prediction accuracy than ESN. As a conclusiong, further investigation will take ESN and RSON as neural network models, and we believe that comparison and evalation works among SRNs, ESNs, and RSONs are also venturing and promising directions.

7. Acknowledgement

8. References

Baddeley, A. & Susan, J. P. (2006). *Working Memory: An Overview. Working Memory and Education*, Academic Press, ISBN:0125544650, Burlington, pp.1-31.

Balota, D. A. (1994). Visual word recognition: the journey from features to meaning, In: *Handbook of Psycholinguistics*, Gernsbacher, M. A., (Ed.), pp.303-358, Academic Press, ISBN-10: 0122808908, New York

Bates, E. & MacWhinney, B. (1987). Competition, variation and language learning, In: MacWhinney, B. (Ed.), *Mechanisms of Language Acquisition*, Erlbaum, Hillsdale, NJ, pp. 157–194.

Berwick, R. & Weinberg, A. (1984). *The Grammatical Basis of Linguistic Performance*, MIT Press, Cambridge, MA.

Bock, K. & Griffin, Z. M. (2000). The persistence of structural priming: Transient activation or implicit learning? *Journal of Experimental Psychology: General*, Vol.129, pp.177–192.

Caplan, D. & Waters, G. S. (1999). Verbal working memory and sentence comprehension, *Behavioral and Brain Sciences*, Vol.22, pp.77–126.

Chang, F.; Bock, K. & Goldberg, A. (2003). Do thematic roles leave traces of their places? *Cognition*, No. 90, pp.29–49.

Chang, F.; Dell, G. S. & Bock, K. (2006). Becoming syntactic, *Psychological Review*, No.113, pp.234–272.

Charniak, E. (1997). Statistical Techniques for Natural Language Parsing, *AI Magazine*, No.18, pp. 33-43.

Chomsky, N. (1959). On Certain Formal Properties of Grammars, *Information and Control*, Vol. 2, pp. 137-167.

Chomsky, N. (2000). *New Horizons in the Study of Language and Mind*, Cambridge University Press, Cambridge, UK.

Culter, A. (1995). Spoke word recognition and production, In: *Speech, Language, and Communication (Handbook of Perception and Cognition)*, Miller, J. L. and Eimas, P. D., (Eds.), pp. 97-136, Academic Press, ISBN-10: 0124977707, New York

Elman, J. L. (1991). Distributed Representations, Simple Recurrent Networks, and Grammatical Structure, *Machine Learning*, Vol. 7, pp. 195-225.

Farkaš , I. & Crocker, M. (2006). Recurrent networks and natural language: exploiting self-organization, in: *Proceedings of the 28th Annual Conference of the Cognitive Science Society*, Lawrence Erlbaum, Hillsdale, NJ.

Ferreira, F. & Clifton, C. (1986). The independence of syntactic processing, *Journal of Memory and Language*, No. 25, pp. 348–368.

Ferreira, V. S. & Bock, K. (2006). The functions of structural priming. Language and Cognitive Processes, 21, 1011–1029.

Fodor, J.A. (1983). *The Modularity of Mind: An Essay on Faculty Psychology*, MIT Press, Cambridge, MA.

Frank, S. (2006). Learn more by training less: systematicity in sentence processing by recurrent networks, *Connection Science*, Vol.18, No.3, pp. 287–302.

Frazier, L. & Fodor, J.D. (1978). The sausage machine: a new two-stage model of the parser, *Cognition*, Vol. 6, pp. 291–325.

Frazier, L. & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences, *Cognitive Psychology*, Vol. 14, pp. 178–210.

Frazier, L. (1987). Sentence processing: a tutorial review, In: Coltheart, M. (Ed.), *Attention and Performance XII*, Erlbaum, London, pp. 559–585.

Friederici, A.D. (2002). Towards a neural basis of auditory sentence processing, *Trends of Cognition Science*, No. 6, pp.78-84.

Friederici, A.D. & Kotz, S.A. (2003). The brain basis of syntactic process: functional imaging and lesion studies, *NeuroImage*, Vol. 20, pp. 8–17.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies, *Cognition*, No. 68, pp. 1–76.

Gómez, R. (2002). Word frequency effects in priming performance in young and older adults, *Journals of Gerontology*, Series B: Psychological Sciences and Social Sciences, 57B(3), pp. 233–240.

Green, T. R. (1979). The necessity of syntax markers: Two experiments with artificial languages, *Journal of Verbal Learning and Verbal Behavior*, Vol. 18, pp.481–496.

Guo, Y. & Shao, Z. (2008). The Cognitive Interactionist Approach of Sentence Parsing with Simple Recurrent Neural Networks, In: *Proceedings of International Symposium on Intelligent Information Technology Application 2008* (IITA 2008), IEEE Computer Society Press, Shanghai, China, pp.118-122.

Hadley, R. (1994). Systematicity in connectionist language learning, *Mind Language*, Vol. 9, No.3, pp. 247–272.

Hauser, M.D.; Chomsky, N. & Fitch, W.T. (2002). The faculty of language: what is it, who has it, and how did it evolve? *Science*, No. 298, pp.1569-1579.

Jackendoff, R. (2002). Foundations of language: brain. Meaning, Grammar, Evolution, *OUP*, Oxford, UK.

Jaeger, H. (2003). Adaptive nonlinear system identification with echo state networks, in: *Advances in Neural Information Processing Systems*, Vol. 15, MIT Press, Cambridge, MA.

Johnson-Laird, P. N. (1983). *Mental Models*, Harvard University Press, ISBN-10: 0674568818, Cambridge, MA.

Just, M. A. & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory, *Psychological Review*, Vol. 99, pp. 122–149.

Kako, E. & Wagner, L. (2001). The semantics of syntactic structures, *Trends in Cognitive Sciences*, Vol. 5, No. 3, pp. 102-118.

Kapetanios, E.; Baer, D. & Groenewoud, P. (2005). Simplifying syntactic and semantic parsing of NL-based queries in advanced application domains, *Data & Knowledge Engineering*, Vol. 55, No.1, pp. 38-58.

Kennison, S. M. (2001). Limitations on the use of verb information in sentence comprehension, *Psychonomic Bulletin and Review*, Vol. 8, pp. 132–138.

Kintsch, W. (1988). The use of knowledge in discourse processing: a construction-integration model, *Psychological Review*, No. 95, pp.163-182, ISSN: 0033-295X

Knight, K.; Chancer, I. & Haines, M. (1995). Filling knowledge gaps in a broad-coverage machine translation system, In : *Proceedings of IJCAI'95*, pp.1390-1397, Montreal, Canada.

Lenat, D.(2006). Hal's Legacy: 2001's Computer as Dream and Reality. From 2001 to 2001: Common Sense and the Mind of HAL, http://www.cyc.com/cyc/technology/halslegacy.html, Retrieved on 2006-09-26.

Lewis, R. L.; Vasishth, S., & VanDyke, J. A. (2006). Computational principles of working memory in sentence comprehension, *Trends in Cognitive Sciences*, Viol. 10, pp. 447–454.

MacDonald, M.C.; Pearlmutter, N.J. & Seidenberg, M.S. (1994). The lexical nature of syntactic ambiguity resolution, *Psychological Review*, No. 101, pp. 676–703.

MacDonald, M. C.; Almor, A. ; Henderson, V. W.; Kempler, D., & Andersen, E. (2001). Assessing working memory and language comprehension in Alzheimer's disease, *Brain & Language*, No. 78, pp. 17–42.

MacDonald, M. C. ; & Christiansen, M. H. (2002). Reassessing working memory: Comment on Just and Carpenter (1992) and Waters and Caplan (1996), *Psychological Review*, No. 109, pp. 35–54.

Marslen-Wislon, W. & Tyler, L.K. (1980). The temporal structure of spoken language understanding, *Cognition*, Vol. 8, pp. 1–71.

Marslen-Wilson, W.D. & Tyler, L.K. (1987). Against modularity, In: Gareld, J.L. (Ed.), *Modularity in Knowledge Representation and Natural-Language Understanding*, MIT Press, Cambridge, MA, pp. 37–62.

McClelland, J.L. ; St John, M. & Taraban, R. (1989). Sentence comprehension: a parallel distributed processing approach, *Language and Cognition Processes*, Vol. 4, pp. 287–336.

Miller, G. A. (1956). The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information, *Psychological Review*, No. 63, pp. 81-97.

Miller, G. A.; Beckwith, R.; Fellbaum, C.; Gross, D., & Miller, K. J. (1990). Introduction to WordNet: An On-Line Lexical Database, *International Journal of Lexicography*, Vol. 3, No. 4, pp. 235-312.

Misyak, J. B. & Christiansen, M. H. (2007). Extending statistical learning farther and further: Long-distance dependencies, and individual differences in statistical learning and language, In: D. S. McNamara & J. G. Trafton (Eds.), *Proceedings of the 29th annual cognitive science society conference*, pp. 1307-1312, Austin, TX: Cognitive Science Society.

Monaghan, P.; Chater, N. & Christiansen, M. H. (2005). The differential role of phonological and distributional cues in grammatical categorisation, *Cognition*, Vol. 96, pp. 143–182.

Morgan, J. L.; Meier, R. P. & Newport, E. L. (1987). Structural packaging in the input to language learning: Contributions of prosodic and morphological marking of phrases to the acquisition of language, *Cognitive Psychology*, Vol. 19, pp. 498–550.

Onnis, L.; Christiansen, M.; Chater, N. & Gómez, R. (2003). Reduction of uncertainty in human sequential learning: Evidence from artificial language learning, In: *Proceedings of the 25th annual conference of the cognitive science society*, Mahwah, NJ: Lawrence Erlbaum, pp. 886–891.

Rayner, K.; Carlson, M., & Frazier, L. (1983). The interaction of syntax and semantics during sentence processing, *Journal of Verbal Learning and Verbal Behavior*, No. 22, pp. 358–374.

Saffran, J. R. (2001). The use of predictive dependencies in language learning, *Journal of Memory and Language*, Vol. 44, No. 4, pp. 493–515.

Saffran, J. R. (2003). Statistical language learning: Mechanisms and constraints, *Current Directions in Psychological Science*, Vol. 12, No. 4, pp. 110–114.

Sedivy, J. C.; Tanenhaus, M. K.; Chambers, C. G. & Carlson, G. N. (1999). Achieving incremental semanticinterpretation through contextual representation, *Cognition*, No. 71, pp. 109–148.

Speer, S. R. & Clifton, C. (1998). Plausibility and argument structure in sentence comprehension, *Memory and Cognition*, Vol. 26, pp. 965–978.

Sugumaran, V. & Storey, V. C. (2002). An ontology based framework for generating and improving DB design, In: *Proceedings of 7th International Workshop on Applications of Natural Language to Information Systems*, pp. 1-12, Springer Verlag, Stockholm, Sweden.

Tanenhaus, M. K.; Spivey-Knowlton, M. J.; Eberhard, K. M. & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension, *Science*, No. 268, pp. 1632–1634.

Tanenhaus, M. K. & Trueswell, J. C. (1995). *Sentence Comprehension. Speech, Language, and Communication (2nd)*, Academic Press, San Diego, pp. 217-262.

Taraban, R. & McClelland, J. R. (1988). Constituent attachment and thematic role assignment in sentence processing: Influence of content-base expectations, *Journal of Memory and Language*, Vol. 27, pp. 597–632.

Townsend, D. J. & Bever, T.G. (2001). *Sentence Comprehension: The Integration of Habits and Rules*, The MIT Press, Cambridge, MA.

Trueswell, J. & Tanenhaus, M. K. (1994). Toward a lexicalist framework for constraint-based syntactic ambiguity resolution, In: C. Clifton, L. Frazier, & K. Rayner (Eds.),

Perspectives on sentence processing, pp. 155–179, Hillsdale, NJ: Lawrence Erlbaum Associates.

Trueswell, J. C.; Tanenhaus, M. K. & Garnsey, S. M. (1994). Semantic Influences On Parsing: Use of Thematic Role Information in Syntactic Ambiguity Resolution, *Journal of Memory and Language*, Vol. 33, No. 3, pp. 285-318.

Trueswell, J. C.; Tanenhaus, M. K. & Kello, C. (1993). Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, Vol. 19, pp. 528–553.

Tulving, E. (1972). Episodic and semantic memory, In : E. Tulving & W. Donaldson (Eds.), *Organization of memory*, New York: Academic Press.

van Assem, M.; Gangemi, A. & Schreiber, G. (2006). *RDF/OWL Representation of WordNet*, W3C Working Draft 19 June 2006, http://www.w3.org/TR/wordnet-rdf/#intrown.

van der Velde, F.; van der Voort, G.; van der Kleij; de Kamps, M.(2004). Lack of combinatorial productivity in language processing with simple recurrent networks, *Connection Science*, Vol. 16, No. 1, pp. 21–46.

van Eijck, J. & Kamp, H. (1997). Representing discourse in context, In: *Handbook of Logic and Language*, van Bentham, J. And ter Meulen, A., (Eds), pp.179-237, Elsevier, ISBN-10: 044481714X, Amsterdam

Zurif, E. B.; Swinney, D.; Prather, P.; WingWeld, A. & Brownell, H. H. (1995). The allocation of memory resources during sentence comprehension: Evidence from the elderly, *Journal of Psycholinguistic Research*, Vol. 24, pp. 165–182.
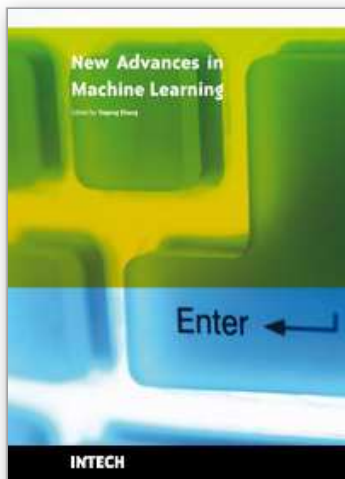
New Advances in Machine Learning

Edited by Yagang Zhang

The purpose of this book is to provide an up-to-date and systematical introduction to the principles and algorithms of machine learning. The definition of learning is broad enough to include most tasks that we commonly call "learning" tasks, as we use the word in daily life. It is also broad enough to encompass computers that improve from experience in quite straightforward ways. The book will be of interest to industrial engineers and scientists as well as academics who wish to pursue machine learning. The book is intended for both graduate and postgraduate students in fields such as computer science, cybernetics, system sciences, engineering, statistics, and social sciences, and as a reference for software professionals and practitioners. The wide scope of the book provides a good introduction to many approaches of machine learning, and it is also the source of useful bibliographical information.

How to reference

In order to correctly reference this scholarly work, feel free to copy and paste the following:

Yi Guo and Zhiqing Shao (2010). Cognitive Learning for Sentence Understanding, New Advances in Machine Learning, Yagang Zhang (Ed.), ISBN: 978-953-307-034-6, InTech, Available from: http://www.intechopen.com/books/new-advances-in-machine-learning/cognitive-learning-for-sentence-understanding

open science | open minds