A Scheme of Word Sense Disambiguation based on Integrated Language Knowledge Base

Jiangsheng Yu* **Shiwen Yu**
yujs@pku.edu.cn yusw@pku.edu.cn
Institute of Computational Linguistics
Peking University, Beijing, China, 100871

Abstract

We intend to provide a broad-brush framework of WSD based on the Integrated Language Knowledge Base (ILKB) in the Institute of Computational Linguistics/Peking University, as the guidelines of our imminent project. A well-structured ILKB contains at least a syntactic lexicon, a semantic lexicon and a large-scale segmented/(POS, concept) tagged corpus, in which the relationship between the method of Computational Lexicology and that of Corpus Linguistics is quite clarified. What's more, the training of concept TagSet along the hypernymy tree is no longer separated from some specific statistical model, such as HMM with two parameters (POS and concept). In short, Statistical Machine Learning will be emphasized in the constructions of both ILKB and TagSet, even throughout WSD project.

1 Introduction

Since the 1980's, Computational Semantics has been playing the keynote in both Natural Language Processing (NLP) and Natural Language Understanding (NLU)[1]. What does it mean that a machine could **understand** a given sentence S or a text T? As we know, Turing Test of NLU includes at least the meaning of any word w in S or T, and as a postulate, WSD, the prerequisite to NLU is to tag the semantic information of w. Undoubtedly, with the development of Machine Translation (MT), Information Extraction (IE), Information Retrieval (IR) and other NLP research domains, WSD has become more and more imperative. [**?**] summarized the development of WSD from 1950's to 1998 in aspects of

1. AI-based methods
   (a) Symbolic methods
   (b) Connectionlist methods
2. Knowledge-based methods
3. Corpus-based methods

[**?**] argued that WordNet, as a computational lexicon, is not a perfect resource for WSD because of the fine-grainedness of the sense distinction. Till now, the automatic construction and the machine learning of a semantic lexicon oriented to some specific applications is still an open problem.[2] Anyway, WordNet[3] in Princeton University provides an approach to the formalization of concepts in natural language, in which a concept is defined by a synonym set (SynSet). A more important work in WordNet is the construction of a well-structured

*Ph.D. of applied mathematics, associate professor, research domains: Computational Semantics, Statistical Linguistics and Computational Lexicology. This paper is supported by National Foundation of Natural Science (Research on Chinese Information Extraction), No. 69483003 and Project 985 in Peking University.

[1]See more details in [**?**], a summarization of WSD.

[2][**?**] studied the evolution of the WordNet-like lexicon, which attempted to strict the automatic training of the semantic descriptions, especially the knowledge structures of the lexicon.

[3]The specification of WordNet could be found in [**?**], [**?**], [**?**], [**?**], [**?**], [**?**], etc.

concept network based on the hypernymy relation (the main framework) and other accessorial relations, such as, the opposite relation, the holonymy relation, etc. The Institute of Computational Linguistics (ICL)/Peking University has been constructing a WordNet-like lexicon named Chinese Concept Dictionary (CCD) since September 2000. With the tool software of Visualized Auxiliary Construction of Lexicon (VACOL), about 20,000 bilingual concepts have been finished. The structure of Knowledge Representation in WordNet-like lexicon is compulsory for WSD, which has been easily verified by its various applications recently ([?], [?], [?], [?], [?], [?] and [?]).

The Integrated Language Knowledge Base in ICL[4], which contains the Grammatical Knowledge Base of Contemporary Chinese (GKB, more than $70,000$ Chinese words with elaborate syntactic descriptions), the Semantic Lexicon oriented to Chinese-English Machine Translation (more than $50,000$ Chinese-English corresponding words with dependency descriptions), CCD and a large-scale segmented/POS tagged corpus (more than $20,000,000$ Chinese characters, the accuracy of segmentation and POS tagging is not less than 99.8%), makes WSD possible as a further application. Based on these resources, the project of WSD include the following six aspects:

1. Segmented/POS and concept tagged corpus checked by hand, which contains $4,000,000$ Chinese characters, as training set

2. Statistical method of WSD, that is, Hidden Markov Model (HMM) with two parameters — Part of Speech (POS) and concept

3. The training of structured concept TagSet for WSD

4. Optimization of HMM

5. Deductive Rules of WSD from ILKB

6. Rule-based and statistics-based approaches to WSD

We just intend to describe a framework of WSD in ICL, in this paper, avoiding to plunge in many details[5]. Although the descriptions are still broad-brush today, the teeming content in WSD makes it attractive for nerved researchers. The main topic in section 2 is the relationship between corpus linguistics and the semantic lexicon, in which we emphasize an empirical construction of concept TagSet. And also, we'll give a precise definition of WSD from the viewpoint of CCD. Section 3 focuses on our primary resolvent of WSD, that is, HMM with two parameters. The topics in the next section are some possible applications of this kind of WSD in NLP products.

2 From Tagged Corpus to Concept TagSet

As we know in Artificial Intelligence, any efficient knowledge representation is determined by an idiographic application. A set of metarules which are able to attemper the existing resources in ICL, especially for WSD, becomes a hot topic currently. After the clarification of the purpose of WSD, ICL will do much research on the construction of the ILKB together with its description framework, on which those metarules work efficiently. For instance, given a concept C, the automatic links to GKB and the segmented/POS tagged corpus by the words in C would certainly benefit the concept tagging by hand as well as the distribution of the concepts in the tagged corpus. Even, we are planning to gather the samples for each concept from the corpus in ICL, which could speed up the construction of the initial concept tagged corpus.

The principle of the more the better is not valid for all databases actually, that is, what we need is a well-structured and efficient corpus, not the largest one. Our main statistical methods for the structure and the size of the ILKB are Pattern Recognition and Interval Estimation respectively. And the Bayesian Statistics, which is good at the analysis of

[4] More details in [?]. One of the advantages of ILKB is the mutual restrictions between distinct tags, which redounds to overcome the data sparseness. And as a methodology, the construction of ILKB is a trend for WSD.

[5] Such as the open problems mentioned in [?]: the role of context, sense division and evaluation of WSD.

small samples, will be tried in the construction of training set. The pioneer experiment of ILKB in ICL, we always believe, will bring about a fresh methodology of NLP.

2.1 Automatic Expansion of Concept Tagged Corpus

An immediate result from the initial training corpus with concept tagging, is the combination expansion by the substitution of synonyms in each SynSet.

Property 2.1 Suppose that the well-formed sentence $S = w_1 w_2 \cdots w_n$ is tagged by the concept sequence $C_1 C_2 \cdots C_n$, then $S' = w'_1 w'_2 \cdots w'_n$ is also well-formed, $\forall w'_i \in C_i$.

On the one hand, the accuracy of expanded corpus ensures a reliable posterior evaluation of the structure of CCD; on the other hand, the sparse data problem of POS tagging will be lightened to the least level without increasing the size of the initial training corpus. It's just like a magnifier of Corpus Linguistics that puts one against ten. This is one of our important motivations to study WSD — expanding the given corpus rationally.

2.2 Definition of WSD

First of all, the aim of WSD must be clarified mathematically.[6] Wittgenstein (or even earlier, Leibniz) believed that the meaning of a given word is nothing but its usages in langue.[7] From the viewpoint of CCD, a concept is just a SynSet, determined by the Principle of Substitution. The word sense of w is $\Delta(w)$, the set of all the SynSets containing w. The map $w \mapsto \Delta(w)$ describes the correspondence between a word and its senses. For instance, all possible senses of the word *tree* are $\{\{\text{tree}\}, \{\text{tree}, \text{tree diagram}\}\}$.

Definition 2.1 For any w in a given text, the process of WSD is to choose a suitable element in $\Delta(w)$ as w's concept tag. Once the sense of w is identified as concept C, machine looks like understanding the meaning of w by the reasonable substitution of its synonyms.

Some deductive rules about SynSets, such as the following property, could be found in ILKB and used in WSD:

Property 2.2 There is no mapping from the set of concepts, Γ, to the set of word senses, Δ. If $C \in \Delta(w_1) \cap \Delta(w_2)$, then $C \in \Gamma$ and $\{w_1, w_2\} \subseteq C$.

The concept TagSet $\mathsf{T} \subset \Gamma$, theoretically, should satisfy that if w is tagged by $C \in \Gamma$, then $\exists! C'(C' \in \Gamma \wedge C \preceq C' \wedge w \in C')$. That is, there is unique path from C to C', the sense of w in the context.

2.3 Training of TagSet related to Statistical Model

The traditional semantic tags are from some ontology, the apriority of which is often criticized by computational linguists. For us, the empirical method must impenetrate each step of WSD because of the complexity of language knowledge. Statistical approach to WSD needs a well concept-tagged corpus as the training set. A corpus with $4,000,000$ Chinese characters is under our consideration as the first step of WSD. Each noun (verb, adjective or adverb) w will be tagged by a specific element in $\Delta(w)$. Because of the sparse data problem, only a real subset of Γ may act as the TagSet of concepts in statistical model, i.e., HMM with two parameters. So, the next step is to find out the satisfiable solution of TagSet. It's easy to see that various corpora could be derived from the original one by the hypernymy relation. Different from those unframed training set, the initial TagSet Γ collapses along the hypernymy trees by some well-defined restrictions, such as

1. avoiding sparse data problem, and
2. the accuracy of some specific statistical model.

The first step leads to a set of structured TagSets $\{\mathsf{T}_1, \mathsf{T}_2, \cdots, \mathsf{T}_n\}$, and the second one is to choose the best one which is the most propitious to the given statistical model. Suppose that $\mathsf{T} = \{C_1, C_2, \cdots, C_m\}$ is the TagSet, and

[6] [?] defines that WSD, at least, involves the association of a given word in a text or discourse with a definition or meaning distinct from other senses potentially attributable to that word.

[7] In [?], Wittgenstein asserted, "Don't look for the meaning, but for the use."

the word w in a given sentence is tagged by C_i, then the meaning of w here is the SynSet C which satisfies that $C_i \preceq C$ and $w \in C$.

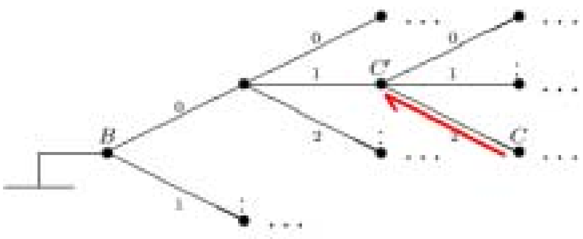

Figure 1: Collapse along the Hypernymy Tree

The general policy of TagSet training is that, TagSet is related to HMM with two parameters and vice versa. By the way, Markov process with two parameters is still a very difficult branch of Random Process Theory even today. Fortunately, the fact that POS and concept are not independent makes the model a little easier, so that POS tagging and concept tagging could benefit from each other. Statistical Machine Learning is emphasized in the constructions of both ILKB and TagSet.

3 Mixed Model for WSD

The most successful model for WSD, imaginatively, is a mixed one, which means that

1. the statistical model is together with a rule-based approach, and
2. the Chinese segmentation, the POS tagging and the concept tagging are unified.

Based on a lexicon, any possible segmented sequence $s = w_1 w_2 \cdots w_n$ corresponds a set of probabilities of POS sequences $A_s = \{\mathsf{P}(P_1^{(i)} P_2^{(i)} \cdots P_n^{(i)})\}$, and each $P_1^{(i)} P_2^{(i)} \cdots P_n^{(i)}$ corresponds a set of probabilities of concept sequences $B_s^{(i)} = \{\mathsf{P}(C_1^{(i,j)} C_2^{(i,j)} \cdots C_n^{(i,j)})\}$, where $C_k^{(i,j)}$ has the POS of $P_k^{(i)}$, then

$$\underset{s}{argmax}(a \cdot \max_s(A_s) + b \cdot \max_{i,s}(B_s^{(i)})) \qquad (1)$$

is the choice, where $a > 0, b > 0$ and $a + b = 1$. The weights of a and b are experiential from the training corpus. The mixed model we understand also includes the rule-based identification of word sense.

3.1 Rule-based WSD

The closed semantic constraints of verb concepts from noun concepts in CCD and the dependency information in the semantic lexicon oriented to Chinese-English Machine Translation will, certainly, be much helpful to WSD. Also, GKB provides some syntactic rules for WSD. The following example shows more persuasion of the semantic constraints:

Example 3.1 The verb *dǎ* has many meanings in Chinese, which differ in *dǎ háizi* (punish the child), *dǎ máoyī* (weaver the sweater), *dǎ jiàngyóu* (buy the soy), etc. In other words, the semantics of arguments determines distinct *dǎ* actually.

Definition 3.1 w is called *unambiguous* if the POS (or concept) of w is unique. w is a *correspondence-unambiguous* (or *c-unambiguous*) word, if there is a bijection between w's POSs and $\Delta(w)$. Otherwise, w is called *correspondence-ambiguous* (or *c-ambiguous*). Obviously, an unambiguous word is correspondence-ambiguous.

By the definition, the verb *dǎ* is c-ambiguous. For those c-unambiguous words, the restriction between POS and concept is simple.

3.2 Statistical Models for WSD

POS and concept tag are two random variables in HMM. Sometimes, POS of w determines a unique SynSet containing w, and sometimes not. But in most cases, a SynSet of w implies a unique POS. The distribution of w's senses with the POS P is also important in the identification of (POS, concept)-tagging in some given context. A Hidden Markov Model with two parameters will be adopted as the main statistical model for WSD, and the Statistical Decision Theory and Bayesian Analysis, which are good at analyzing the small samples, conducted as a comparison. The default initial training corpus T is a subset of the well segmented/POS tagged People's Daily in ICL. The pretreatment of T refers to the noise filtering and concept tagging by hand. Some convenient software will be provided, such as the cursor sensitive for the automatic display and choosing of the word senses.

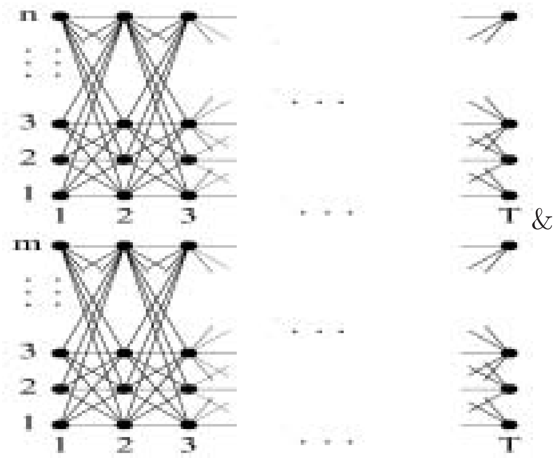

Figure 2: HMM of POS and concept

Definition 3.2 For the result of automatic segmentation and POS tagging by the lexicon $s = w_1/_{P_1^{(i)}}\ w_2/_{P_2^{(i)}} \cdots w_n/_{P_n^{(i)}}$, define

$$f(i) = \underset{j}{argmax}\, \mathsf{P}(C_1^{(i,j)} \cdots C_n^{(i,j)} | P_1^{(i)} \cdots P_n^{(i)}) \quad (2)$$

Thus, there is a map g from the set of $\{P_1^{(i)} P_2^{(i)} \cdots P_n^{(i)}\}$ to the set of $\{C_1^{(i,j)} C_2^{(i,j)} \cdots C_n^{(i,j)}\}$, which satisfies that

$$g(P_1^{(i)} \cdots P_n^{(i)}) = C_1^{(i,f(i))} \cdots C_n^{(i,f(i))} \quad (3)$$

where $\forall i, k, \exists C \in \Delta(w_k)$ s.t. $C_k^{(i,f(i))} \preceq C$. If there is $C' \neq C$ satisfying $C' \in \Delta(w_k)$ and $C_k^{(i,f(i))} \preceq C'$, then the one with more distribution is the selected sense of w_k with the POS of $P_k^{(i)}$ in the context. Nevertheless, this is just one result of the segmentation and POS tagging by lexicon. A method which is more efficient than that described in equation-**??**, is

$$\underset{s}{argmax}\{\max_i\{a \cdot \mathsf{P}(P_s^{(i)}) + b \cdot \mathsf{P}(g(P_s^{(i)}))\}\} \quad (4)$$

where $P_s^{(i)} = P_1^{(i)} P_2^{(i)} \cdots P_n^{(i)}$ and a, b are experiential weights.

3.3 Optimization of HMM

Animated by Baum-Welch algorithm of the HMM with one parameter, the optimization of statistical model for WSD leads the data to a satisfiable state, which is in some sense a dynamic design of HMM. Not only the training of large scale data set in a derived corpus, but also the comparison between any two possible well-defined TagSets in a closed testing obviously effect the accuracy of concept tagging, in which the distribution of word senses also plays an important role.

3.4 Congruity between Statistical and Rule-based Approaches

The principle is that except certain cases in which the rule-based approach works, the statistical model acts as the protagonist in WSD. An important research is to find out the distribution of mistakes caused by the statistical model firstly, and then to study whether they could be corrected by some rules type by type. The worst situation is that there is no method that could be used to keep an optimum for a class of words, phrases or sentences. The statistical distributions of these obstinate types are also valuable for us because that if such kinds of noises are deleted from the initial training corpus, the statistical data in HMM will be more accurate.

4 NLP Benefits from WSD

WSD is, in fact always, the kernel problem of both NLU and NLP ([**?**], [**?**]), which could be so straightforwardly applied to many other NLP products. Any development of the scheme for WSD brings about a revolution in Computational Linguistics — it is the prelude of semantic analysis of natural languages. With the implementation of WSD step by step, the existing software tools of IE, IR, MT will be improved accordingly. And the language resources, as concomitants, become the necessities of further applications. As a static lexicon, WordNet just represents a single ontology that the psycholinguists are interested in. Actually, the structure of the so-called common knowledge is nothing but a statistical distribution, which is effected by the cultures and personal experiences. Oriented to a specific application, such as IE, the appropriate information in a WordNet-like lexicon seems necessary.

Example 4.1 $C = \{$*earthquake, quake, temblor, seism*$\}$ is not only a kind of $C' = \{$*geological phenomenon*$\}$, but also a kind of $C'' = \{$*natural disaster*$\}$. Suppose that there are n angles of view, $v_1, v_2, \cdots, v_n$, then the hypernymy relation could be classified into $\{h_1, h_2, \cdots, h_n\}$, where h_i is defined by v_i. A geographer prefers $C' \prec_{v_i} C$, while the common people think $C'' \prec_{v_j} C$ more reasonable.

4.1 Industry Development

The explosion of information in the network conduces to the development of NLP technologies. Applied systems of IE, IR and MT are really the urgent needs nowadays. However, almost all existing software tools of NLP related to semantics have the problem of WSD. For instance, word matching in search engine system could guarantee neither a satisfiable precision nor a satisfiable recall ratio due to the lack of concept identification. The freedom of looking for the interesting topics in thousands of rubbish results detected by the simple word-matching algorithm is not what we want. If the pretreatment involves WSD in text classification and shallow parsing, then billions of people will benefit from the reformative research results of WSD. Nothing could be more far-reaching than WSD in NLP. In other words, the commercial value of WSD today is that of NLP in the future.

4.2 Technology Capability Building

The focus of this project is the interdisciplinary study of WSD in Statistics-based and rule-based approaches. Unifying the program is an emphasis on the sequential resources for different purposes, in which the ultimate outcome of a segmented/POS and concept tagged corpus by machine could assess not only the validity of HMM but also the efficiency of TagSet training. It's a completely empirical method of WSD that reflects the current development of NLP and Computational Linguistics typically. We believe that this framework applies to a great many NLP problems faced by both linguists and computational linguists, and that the use of a common framework will deepen theoretical insights into the nature of NLU. It means that the clarification of the classified difficulties in WSD and the limitation of statistical methods, in the nature of things, will redound to the new technologies of semantic analysis in Computational Linguistics. At least in this project, technologies of the HMM with two parameters and its optimization, shallow parsing of chunk sequence, construction of language knowledge base and semantic analysis, will be developed mostly.

Conclusion

We described a scheme of WSD based on ILKB, as a groping approach to the statistical concept tagging. The empirical method is emphasized in each step of WSD, such as TagSet training and HMM with two parameters. And we notice that the construction of ILKB should be oriented to its specific applications, that is, the machine learning of lexicon is also necessary in WSD.[8] The recent research is the simulative experiment of TagSet training and HMM with two parameters.

Acknowledgement

We appreciate all the colleagues participating in ILKB project, especially Prof. ZHU Xuefeng, Prof. DUAN Huiming, Prof. LU Chuan, Dr. Zhan Weidong, etc. The blithesome collaboration with the members of CCD group is memorable for all of us. Lastly, many thanks to our friends in the Second Workshop on Chinese Lexical Semantics for their kindly discussion with the authors.

References

[1] Ahlswede T. E. 1995 *Word Sense Disambiguation by Human Informants*, in Proceedings of the Sixth Midwest Artificial Intelligence and Cognitive Society Conference, Carbondale, Illinois, April 1995, pp.73-78.

[2] ALPAC 1966 *Language and Machine: Computers in Translation and Linguistics*, National Research Council Automatic Language Processing Advisory Committee, Washington, D.C.

[3] Alshawi H. and Carter D. 1994 *Training and Scaling Preference Functions for Dis-*

[8] In [?], the evolution of WordNet-like lexicon, especially the knowledge structure, is studied.

ambiguation, in Computational Linguistics, 20(4), pp635-648.

[4] Aristotle 1941 *Categoriae*, in The Basic Works of Aristotle, R. McKeon (ed). Random House, New York.

[5] Basili R. and Della R. et al 1997 *Towards a Bootstrapping Framework for Corpus Semantic Tagging*, in ACL-SIGLEX Workshop *Tagging Text with Lexical Semantics: Why, What, and How?*, pp66-73.

[6] Beckwith R. 1998 *Design and Implementation of the WordNet Lexical Database and Searching Software*, in [?], pp105-127.

[7] Bruce R. and Wiebe J. 1994 *Word-sense Disambiguation Using Decomposable Models*, in Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics, Las Cruces, New Mexico, pp139-145.

[8] Carnap R. 1966 *Der Logische Aufbau Der Welt*. Felix Meiner Verlag, Hamburg.

[9] Carpenter B. 1992 *The Logic of Typed Feature Structures*. Cambridge University Press.

[10] Fellbaum C. 1998 *A Semantic Net of English Verbs*, in [?], pp69-104.

[11] Fellbaum C. (ed) 1999 *WordNet: An Electronic Lexical Database*, The MIT Press.

[12] Huang C.R. et al 2001 *Linguistic Tests for Chinese Lexical Semantic Relations: Methodology and Implications*, report in the Second Workshop on Chinese Lexical Semantics.

[13] Ide N. and Véronis J. 1998 *Introduction to Special Issue on Word Sense Disambiguation: The State of Art*, Computational Linguistics, Vol. 24, No. 1, pp1-40.

[14] Krenn B. and Samuelsson C. 1997 *The Linguist's Guide to Statistics*.

[15] Lakoff G. 1987 *Women, Fire, and Dangerous Things: What Categories Reveal about the Mind*. University of Chicago Press.

[16] Leacock C. and Chodorow M. 1998 *Combining Local Context and WordNet Similarity for Word Sense Identification*, in [?], pp265-284.

[17] Leibniz G.W. 1981 *New Essays on Human Understanding*, P. Remnant and J. Bennett (Ed. and trans.), Cambridge University Press.

[18] Lesk M. 1986 *Automatic Sense Disambiguation Using Machine Readable Dictionaries: How to Tell a Pine Cone from an Ice Cream Cone*, Proc. of SIGDOC, Toronto, pp1-9.

[19] Liddy E. D. and Paik W. 1993 *Statistically-guided Word Sense Disambiguation*, in Proceedings of the AAAI Fall Symposium Series, pp98-107.

[20] Lucas W.F. (ed) 1983 *Modules in Applied Mathematics*. Spring-Verlag New York, Inc.

[21] McRoy S. W. 1992 *Using multiple knowledge Sources for Word Sense Discrimination*, in Computational Linguistics, 18(1), pp1-30.

[22] Miller G.A. et al 1993 *Introduction to WordNet: An On-line Lexical Database*, in the attached specification of WordNet 1.6.

[23] Miller G.A. 1998 *Nouns in WordNet*, in [?], pp23-46.

[24] Partee B.H. et al 1990 *Mathematical Methods in Linguistics*. Kluwer Academic Publishers.

[25] Priss U. 1999 *The Formalization of WordNet by Methods of Relational Concept Analysis*, in [?], pp179-196.

[26] Quine W. 1980 *From a Logical Point of View*. Harvard University Press.

[27] Rosch E. 1975 *Cognitive Representations of Semantic Categories*. Journal of Experimental Psychology, 104, 192-233.

[28] Russell B. 1948 *Human Knowledge — Its Scope and Limits*. Simon and Schuster.

[29] Russell B. 1989 *Logic and Knowledge*. Unwin Hyman Ltd.

[30] Sanderson M. 1994 *Word Sense Disambiguation and Information Retrieval* in Croft, W. and van Rijsbergen, C. J. (Eds.), Proceedings of the 17th Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval, Las Vegas, pp161-175.

[31] Simpson G. B. and Burgess C. 1989 *Implications of Lexical Ambiguity Resolution for Word Recognition and Comprehension*, in Small S. and Cottrell G. et al (Eds.) 1988 *Lexical Ambiguity Resolution: Perspectives from Psycholinguistics, Neuropsychology, and Artificial Intelligence*, Morgan Kaufman, San Mateo, California, pp271-288.

[32] Slator B. M. 1992 *Sense and Preference*, in Computer and Mathematics with Applications, 23(6/9), pp391-402.

[33] Sussna M. 1993 *Word Sense Disambiguation for Free-text Indexing Using a Massive Semantic Network*, in Proceedings of the Second International Conference on Information and Knowledge Base Management, CIKM'93, Arlington, Virginia, pp67-74.

[34] Véronis J. and Ide N. 1990 *Word Sense Disambiguation with Very Large Neural Networks Extracted from Machine Readable Dictionaries*, Proc. of COLING-90, Helsinki, vol. 2, pp389-394.

[35] Vossen P. (ed.) 1998 *EuroWordNet: A Multilinugual Database with Lexical Semantic Networks*. Dordrecht: Kluwer.

[36] Wittgenstein L. 1953 *Philosophical Investigations*. Basil Blackwell Ltd.

[37] Yu J.S. 2001 *Structures in CCD*, report of Second Workshop on Chinese Lexical Semantics, held in Peking University.

[38] Yu J.S. 2001 *Algebraic Structures in Linguistics*, report of ICL-salon of Computational Linguistics, Peking University.

[39] Yu J.S. and Yu S.W. et al 2001 *Introduction to Chinese Concept Dictionary*, in International Conference on Chinese Computing (ICCC2001), pp361-367.

[40] Yu J.S. 2000 *Hidden Markov Model and its Applications in NLP*, report of ICL Seminar of Natural Language Processing, Peking University.

[41] Yu J.S. 2000 *Machine Segmentation Ambiguities and Dynamic Lexicon*, in Associated Conference AI2000.

[42] Yu J.S. 2001 *Construction of Semantic Lexicon*, draft in the author's homepage http://icl.pku.edu.cn/yujs

[43] Yu J.S. 2001 *Specification of CCD*, draft of ICL, Peking University, 2000.

[44] Yu J.S. 2001 *The Structure of Chinese Concept Dictionary*, accepted by Journal of Chinese Information Processing, 2001.

[45] Yu J.S. 2001 *Evolution of WordNet-like Lexicon*, accepted by the First WordNet Conference, Mysore, India, 2002.

[46] Yu S.W. 2000 *The Comprehensive Chinese Language Knowledge Base and its Applications in the Teaching of Chinese Language*, in the 4th Global Chinese Conference on Computers in Education.