

WSD and Closed Semantic Constraint

Jiangsheng Yu*

Institute of Computational Linguistics
Peking University, Beijing, China, 100871

Abstract *The application-driven construction of lexicon has been emphasized as a methodology of Computational Lexicology recently. We focus on the closed semantic constraint of the argument(s) of any verb concept by the noun concepts in a WordNet-like lexicon, which theoretically is related to Word Sense Disambiguation (WSD) at different levels. From the viewpoint of Dynamic Lexicon, WSD provides a way of automatic construction for the closed semantic constraints and also benefits from the semantic descriptions.*

Keywords *dynamic lexicon, evolution, WSD, WordNet-like lexicon, closed semantic constraint*

1 Introduction

As the underlying resource of semantic analysis, the most important descriptions in a semantic lexicon are the relationships between verbs and nouns, which usually comes down to the closed semantic constraint of each argument of any verb. Once the structure of the semantic lexicon is determined, the closed semantic constraint becomes a well-defined problem.

Example 1.1 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. Actually, the semantics of *dǎ* is distinguished by the semantics of its arguments.

Different from the traditional lexicon, we advocate the conception of dynamic lexicon ([15]) and its evolution oriented to some particular application, which will be mentioned in Section 2. The WordNet-like lexicon is treated as a dynamic one, which means that the structures representing semantic knowledge could be changed according to some empirical standards. In the next section, we'll define the WSD based on the WordNet-like lexicon, and then discuss the training of concept TagSet and the statistical model of WSD. By the statistical WSD, in Section 4, we introduce an approach to the automatic construction of the closed

semantic constraints in a WordNet-like lexicon. The last section is the conclusion.

2 Dynamic Lexicon and Its Structural Evolution

Definition 2.1 A *dynamic lexicon* is a triple $\text{Lex} = \langle S, R, T \rangle$ in which

1. S is a well-structured set with a type¹ t ,
2. R is the set of deductive rules on S , and
3. T is the set of all structural transformations of S , keeping the type t .

Definition 2.2 Lexicon $\langle S', R, T \rangle$ is called the *evolution result* of the lexicon $\langle S, R, T \rangle$ if $\exists t \in T^*$ such that $S \xrightarrow{t} S'$ (or briefly $S \rightsquigarrow S'$). The process of a dynamic lexicon to its evolution result is called an *evolution*. Obviously, T is a group with the operation of composition.

Definition 2.3 $\langle S, R, T \rangle$ is called *simple structured* if T is a commutative group, otherwise *complex structured*.

The more complex is the structure of S , the more difficult are the applications of R and T . Since some part of semantic knowledge is represented by the structure, the complexity balance between the structure and R (or T) is one of the serious problems in Computational Lexicology.

Definition 2.4 Let $\Omega(S)$ denote the least number of operations constructing S , and $\Omega(S \rightsquigarrow S')$ the least number of operations from S to S' . It's easy to verify that

Theorem 2.1 $\Omega(\cdot)$ is a distance, i.e., it satisfies that $\forall S, S', S''$,

1. $\Omega(S \rightsquigarrow S') \geq 0$
2. $\Omega(S \rightsquigarrow S') = 0 \Leftrightarrow S = S'$
3. $\Omega(S \rightsquigarrow S') = \Omega(S' \rightsquigarrow S)$
4. $\Omega(S \rightsquigarrow S'') \leq \Omega(S \rightsquigarrow S') + \Omega(S' \rightsquigarrow S'')$

*This paper is supported by National Foundation of Natural Science (Research on Chinese Information Extraction), No. 69483003 and Project 985 in Peking University.

¹For instance, labeled tree or complete lattice.

Corollary 2.1 $\Omega(S') \leq \Omega(S) + \Omega(S \rightsquigarrow S')$

Definition 2.5 The *degree of structural destruction* from S to S' , is defined by

$$\rho(S \rightsquigarrow S') = 1 - \frac{\Omega(S')}{\Omega(S) + \Omega(S \rightsquigarrow S')} \quad (1)$$

Property 2.1 $0 \leq \rho(S \rightsquigarrow S') \leq 1$

Definition 2.6 Let $S \rightsquigarrow S_1 \rightsquigarrow \dots \rightsquigarrow S_n \rightsquigarrow \dots$ be a sequence of evolution, the sequence is called *convergent* if there exists a constant A s.t. $0 \leq A \leq 1$ and $\lim_{n \rightarrow \infty} \rho(S \rightsquigarrow S_n) = A$.

It's easy to see that a local evolution of the lexicon may not be an optimization even for a specific application. The index ρ indicates the convergence of lexical structure, guaranteeing a stable machine learning of the dynamic lexicon. 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 particular application, such as IE, IR, MT, etc, the appropriate semantic descriptions in a WordNet-like lexicon seem necessary.

Example 2.1 $C = \{\text{earthquake, quake, tremblor, seism}\}$ is not only a kind of $C' = \{\text{geological phenomenon}\}$, but also a kind of $C'' = \{\text{natural disaster}\}$.

3 WSD based on WordNet-like Lexicon

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 . Thus, the prerequisite WSD is to tag the semantic information of w automatically. WordNet² in Princeton University, in despite of its disputed quality, 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 concept network based on the hypernymy relation (the main framework) and other accessorial relations, such as, the opposite relation, the holonymy relation, entailment, cause, etc.

Definition 3.1 A *WordNet-like lexicon* is a dynamic lexicon with the type of WordNet:

1. restricted to each category, S is a labeled tree from the viewpoint of the hypernymy relation for both noun concepts and verb concepts,

2. some accessorial relations between the noun (or verb) concepts, and
3. closed semantic constraint of the argument(s) of each verb concept from the noun concepts.

The WordNet-like lexicon is complex structured, it may not have the same ontology of WordNet, neither the semantic knowledge representations. But the description method seems a general format for all languages from the fact of EuroWordNet (see [12]), Chinese Concept Dictionary (CCD, see [7], [13], [14] and [15]), Korean WordNet, Tamil WordNet, etc.

Definition 3.2 Let Σ be the set of all words, then Γ , the set of all concepts (or SynSets) in a WordNet-like lexicon, is a subset of 2^Σ . The set of all SynSets containing w is denoted by $\Delta(w)$, in which each element is called a *sense* of w .

Definition 3.3 Given a well-defined sentence $S = w_1 w_2 \dots w_n$, WSD is the computable processing which tags w_i a unique sense $s_i = \{w_i, w_{i_1}, \dots, w_{i_k}\}$ such that each derived combinatorial path is a well-defined sentence with the semantics of S . The Principle of Substitution provides a corpus-based empirical approach to test a SynSet well-defined or not. The SynSet is the smallest unit in a WordNet-like lexicon, which is the underlying of the structural descriptions between the concepts.

The training of concept TagSet and the statistical model of WSD are interactional, which is the main idea of our approach to WSD based on a WordNet-like lexicon.

3.1 The Training of TagSet

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. The statistical approach to WSD needs a well concept-tagged corpus as the training set for the concept TagSet and the statistical data in the Hidden Markov Model (HMM). To avoid the sparse data problem, only a few real subsets of Γ could act as the TagSet in the statistical model (see [15] and [16]). The first step leads to a set of structured TagSets $\{\mathbb{T}_1, \mathbb{T}_2, \dots, \mathbb{T}_m\}$, then the second step is to choose the most efficient one which makes the best accuracy of the statistical concept tagging. Different from those unframed tags, the deductive rule along the hypernymy trees works out the sense of w by the following property:

²The specification of WordNet could be found in [3], [4], [5], [9], [10], [11], etc.

Property 3.1 Suppose that the TagSet is $\mathbf{T} = \{C_1, C_2, \dots, C_k\}$, and the word w in a given sentence is tagged by C_i , then the sense of w here is the SynSet C which satisfies that $C_i \preceq C$ and $w \in C$, where \preceq is the partial ordering of the nodes in the hypernymy tree.

3.2 Statistical Model of WSD

In some sense, WSD is the kernel problem of both NLU and NLP ([1], [6], [8]). POS and concept tag are two random variables in the HMM of WSD. Sometimes POS of w determines its sense, sometimes not. But in most cases, a sense of w implies a unique POS. The distribution of w 's senses with the POS, P , is important in the (POS, concept)-tagging. 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 training corpus, T , is done by hand, where the cursor sensitive display of the senses provides the help information.

Definition 3.4 Consider the well-defined sentence $S = w_1 w_2 \dots w_n$. By the lexicon, let $S = w_1/P_1^{(i)} w_2/P_2^{(i)} \dots w_n/P_n^{(i)}$ be a possible POS tagged result, where $i \in I$. Define

$$\begin{aligned} f(i) &= \underset{j \in J}{\operatorname{argmax}} \mathbb{P}(C_1^{(i,j)} \dots C_n^{(i,j)} | P_1^{(i)} \dots P_n^{(i)}) \\ &= \underset{j \in J}{\operatorname{argmax}} \mathbb{P}(C_1^{(i,j)} \dots C_n^{(i,j)}, P_1^{(i)} \dots P_n^{(i)}) \quad (2) \\ &= \underset{j \in J}{\operatorname{argmax}} \mathbb{P}(C_1^{(i,j)} \dots C_n^{(i,j)}) \end{aligned}$$

The HMM of concept can simulate the HMM with two parameters of (POS, concept). $f(i)$ in (2) is predigested to

$$f(i) = \underset{j \in J}{\operatorname{argmax}} \mathbb{P}(C_1^{(i,j)}) \prod_{k=2}^n \mathbb{P}(C_k^{(i,j)} | C_{k-1}^{(i,j)}) \quad (3)$$

Property 3.2 There exists a unique map g from the set of $\{P_1^{(i)} P_2^{(i)} \dots P_n^{(i)} | i \in I\}$ to the set of $\{C_1^{(i,j)} C_2^{(i,j)} \dots C_n^{(i,j)} | (i, j) \in I \times J\}$, which satisfies that

$$g(P_1^{(i)} \dots P_n^{(i)}) = C_1^{(i,f(i))} \dots C_n^{(i,f(i))} \quad (4)$$

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 .

Property 3.3 Let $s = w_1 w_2 \dots w_n$ be any possible segmented sequence of S , corresponding a set of probabilities of POS sequences $A_s = \{\mathbb{P}(P_1^{(i)} P_2^{(i)} \dots P_n^{(i)}) | i \in I\}$. Each $P_1^{(i)} P_2^{(i)} \dots P_n^{(i)}$

corresponds a set of probabilities of concept sequences $B_s^{(i)} = \{\mathbb{P}(C_1^{(i,j)} C_2^{(i,j)} \dots C_n^{(i,j)}) | j \in J\}$, where $C_k^{(i,j)}$ has the POS of $P_k^{(i)}$, then

$$\underset{s}{\operatorname{argmax}} (a \cdot \max_s(A_s) + b \cdot \max_{i,s}(B_s^{(i)})) \quad (5)$$

is the choice of segmentation, where $a > 0, b > 0$ and $a + b = 1$. More precisely, (5) is rewritten by

$$\underset{s}{\operatorname{argmax}} \{ \max_i \{ a \cdot \mathbb{P}(P_s^{(i)}) + b \cdot \mathbb{P}(g(P_s^{(i)})) \} \} \quad (6)$$

where $P_s^{(i)} = P_1^{(i)} P_2^{(i)} \dots P_n^{(i)}$.

4 WSD driven Closed Semantic Constraint

From the corpus and the statistical WSD, we can make an induction of the arguments along the hypernymy tree, which leads to the closed semantic constraints automatically. At the same time, the closed semantic constraints also provide a possible approach to the empirical optimization of Γ_N and Γ_V . While the total optimization of a WordNet-like lexicon is still an open problem.

4.1 Similarity between Concepts

Definition 4.1 A *labeled tree* is a 5-tuple $T = \langle N, Q, D, P, L \rangle$ satisfying that:

1. N is a finite set of nodes
2. Q is a finite set of labels
3. D is a partial ordering on N , called *dominance relation*
4. P is a strict partial ordering on N , called *precedence relation*
5. $(\exists x \in N)(\forall y \in N)[(x, y) \in D]$
6. $(\forall x, y \in N)[[(x, y) \in P \vee (y, x) \in P] \leftrightarrow [(x, y) \notin D \wedge (y, x) \notin D]]$
7. $(\forall x, y, z, w \in N)[[(w, x) \in P \wedge (w, y) \in D \wedge (x, z) \in D] \rightarrow (y, z) \in P]$
8. $L : N \rightarrow Q$ is a *label map*

Definition 4.2 A hypernymy tree is a labeled tree, in which the label map is one-to-one. Always, we presume that the hypernymy tree is not degenerative.

In a hypernymy tree of a WordNet-like lexicon, a node is a code and a label is a SynSet. Since the label map is injective, without generality, a SynSet is usually denoted by a node. We assume that the precedence relation between

the brother nodes always implies an ordering of time, usage, frequency, mood, etc. For instance, $\{\text{spring, springtime}\} \prec \{\text{summer, summertime}\} \prec \{\text{fall, autumn}\} \prec \{\text{winter, wintertime}\}$ as the hyponyms of $\{\text{season, time of year}\}$.

Definition 4.3 Let f, b and B denote *father, the nearest younger-brother* and *the nearest elder-brother* respectively, satisfying that $f = fb, f = fB$ and $Bb = bB = \mathbf{1}$.

Definition 4.4 $\forall x, y \in N$, let $z \in N$ be their nearest ancestor satisfying $z = f^m(x)$ and $z = f^n(y), D(x, y) \stackrel{\text{def}}{=} m + n$. $k \in \mathbb{N}$ is called the *offset* of x from its eldest brother if $\exists B^k(x)$ and $\nexists B^{k+1}(x)$. Let the offset of y is l , the *similarity* between x and y is:

- If $mn = 1, S(x, y) \stackrel{\text{def}}{=} \langle 0, |k - l| \rangle$
- If $mn \neq 1, S(x, y) \stackrel{\text{def}}{=} \langle m + n, 0 \rangle$

Definition 4.5 Suppose that $S(x_1, y_1) = \langle a_1, b_1 \rangle$ and $S(x_2, y_2) = \langle a_2, b_2 \rangle$, the comparison of similarities is defined as follows:

1. $a_1 = a_2$
 - $S(x_1, y_1) \preceq S(x_2, y_2) \leftrightarrow b_1 \leq b_2$
2. $a_1 \neq a_2$
 - If $a_1 < a_2$, then $S(x_1, y_1) \prec S(x_2, y_2)$
 - If $a_1 > a_2$, then $S(x_2, y_2) \prec S(x_1, y_1)$

Theorem 4.1 $\langle \{S(x, y) | x, y \in N\}, \preceq \rangle$ is a totally ordered set.

The elementary structural transformations in a WordNet-like lexicon include:

1. *insert* a non-root brother-node;
2. *collapse* a non-root node to its father-node;
3. *root* is adding a new root;
4. *add* a link between two labeled trees;
5. *delete* a link between two labeled trees.

4.2 Induction of Constraints

Γ_N (or Γ_V) denotes the set of noun (or verb) concepts. Let $C \in \Gamma_V$ be a verb concept with one argument. Suppose that we have gotten the initial closed semantic constraint of its argument, $C' \in \Gamma_N$, from a concept-tagged sentence. A link from C' to C is added between Γ_N and Γ_V . If C'' from another sentence is also a close semantic constraint of C 's argument, then the infimum of C' and C'' , $\inf(C', C'')$, is the new C' . $\forall x \in \Gamma, C' \preceq x$, if the substitution from C to x still induces well-formed sentences, then the induction succeeds. Otherwise, the disjointed union $C' \oplus C''$ is the closed semantic constraint.

Definition 4.6 The induction of the closed semantic constraints of $C, D \in \Gamma$ is defined by

$$C \sqcap D = \begin{cases} \inf(C, D) & \text{if } \forall x[\inf(C, D) \preceq x] \\ & \text{succeeds in the substitution} \\ C \oplus D & \text{otherwise} \end{cases}$$

Definition 4.7 By **Theorem 4.1**, the induction between $C \oplus D$ and $E \in \Gamma$ is defined by

$$(C \oplus D) \sqcap E = \begin{cases} (C \sqcap E) \oplus D & \text{if } S(C, E) \preceq S(D, E) \\ C \oplus (D \sqcap E) & \text{otherwise} \end{cases}$$

Theoretically, if $C_1 \oplus C_2 \oplus \dots \oplus C_n$ is the closed semantic constraint of the argument of $C \in \Gamma_V$, then $\forall i, \forall x[C_i \preceq x]$ succeeds in the substitution. Thus, in the WordNet-like lexicon, there are n links from Γ_N to Γ_V for C , where n is called the *length of the constraint*. The approach to the closed semantic constraints of the verb concepts with two arguments is similar.

4.3 Clustering of Constraints

Definition 4.8 Suppose that there are N arguments for all verb concepts and the length of the i -th constraint is l_i , then $\bar{l} = \sum_{i=1}^N l_i / N$ is called the *average length of the constraints*.

\bar{l} indicates the rationality of the concept classification in a WordNet-like lexicon, which also acts as an index of the evolution. Our presupposition is that the optimization of the lexicon must have the least average length of the constraints. The clustering of noun concepts constrained by the arguments of the verb concepts should be a standard of the classification of Γ_N .

Definition 4.9 $S \rightsquigarrow S_1 \rightsquigarrow \dots \rightsquigarrow S_n \rightsquigarrow \dots$ is *Cauchy sequence of evolution* iff $\forall \epsilon > 0, \exists N \in \mathbb{N}, \forall i, j > N, \rho(S_i \rightsquigarrow S_j) < \epsilon$.

Theorem 4.2 The Cauchy sequence of evolution is convergent. And $\forall \epsilon > 0, \exists i, j \in \mathbb{N}$ s.t. $|\bar{l}(S_i) - \bar{l}(S_j)| < \epsilon$.

Γ_N is structured by not only the hypernymy relation but also the closed semantic constraints. Of course, the hypernymy relation in Γ_N is principal, but not necessarily unique. As described in **Example 2.1**, the distinct angles of view provide enough space for the evolution. By the hypernymy relation in Γ_V , we have

Property 4.1 $\forall C, C' \in \Gamma_V, C \preceq C'$, if the closed semantic constraint of C' is $C_1 \oplus C_2 \oplus \dots \oplus C_n$, then $\exists C_{n+1}, \dots, C_m \in \Gamma_N$ such that $((C_1 \oplus C_2 \oplus \dots \oplus C_n) \sqcap C_{n+1}) \sqcap \dots \sqcap C_m$ is the closed semantic constraint of C .

This property provides an approach to the empirical testing of the concept classification of Γ_V if Γ_N is fixed. Separately, Γ_N (or Γ_V) can be evaluated by some indexes and evolves to a satisfiable result. A little more complicated, the closed semantic constraints destroy the independent evolution of Γ_N and Γ_V . If Γ_V is fixed, then the optimization of Γ_N may be implemented (but not completely reliable) and vice versa. While it is still an open problem to define a numerical measure that could formalize the optimization of the total structures in a WordNet-like lexicon, especially Γ_N and Γ_V .

5 Conclusion

A scheme of the closed semantic constraint in a WordNet-like lexicon based on WSD has been described as an application driven construction of a dynamic lexicon. At the same time, the further topic leads to how the rule-based concept tagging benefits from the descriptions of semantic constraint. The empirical method is much emphasized in the WSD and the development of the dynamic lexicon, such as the TagSet training, the SynSet testing and the evolution of a WordNet-like lexicon (see [15], [16] and [17]). The author believes that the computable part in Computational Lexicology is nothing but the evolution of the dynamic lexicon oriented to a particular application, which is actually the optimization of the language knowledge base.

Acknowledgement

I appreciate all my colleagues participating in the CCD project, the blithesome collaboration with them is always memorable for me. Many thanks to my friends in the Second and the Third Workshop on Chinese Lexical Semantics for their kindly discussion with the author. Lastly, the most thankful words are given to my wife for her longtime tolerance to my weaselling from the housework under the false pretense of research.

References

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

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

[3] Beckwith R. 1998 *Design and Implementation of the WordNet Lexical Database and Searching Software*, in [5], pp105-127.

[4] Fellbaum C. 1998 *A Semantic Net of English Verbs*, in [5], pp69-104.

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

[6] 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.

[7] Liu Y, Yu S.W. and Yu J.S. *Building a Bilingual WordNet: New Approaches and algorithms*, accepted by COLING2002.

[8] Manning C.D. and Schütze H. 1999, *Foundations of Statistical Natural Language Processing*, The MIT Press.

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

[10] Miller G.A. 1998 *Nouns in WordNet*, in [5], pp23-46.

[11] Priss U. 1999 *The Formalization of WordNet by Methods of Relational Concept Analysis*, in [5], pp179-196.

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

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

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

[15] Yu J.S. 2001 *Evolution of WordNet-like Lexicon*, in The First International Conference of Global WordNet, Mysore, India, 2002.

[16] Yu J.S. and Yu S.W. 2002 *Word Sense Disambiguation based on Integrated Language Knowledge Base*, in The 2nd International Conference on East-Asian Language Processing and Internet Information Technology (EALPIIT'2002).

[17] Yu J.S. 2002 *Statistical Methods in Word Sense Disambiguation*, draft (can be downloaded from <http://icl.pku.edu.cn/yuj/s/>) of seminar at the Institute of Computational Linguistics, Peking Univ..