

# Parsing the Penn Chinese Treebank with Semantic Knowledge

---

Deyi Xiong et al.

IJCNLP 05

*Institute of Computing Technology, Chinese Academy  
of Sciences*

<http://mtgroup.ict.ac.cn/~devi/>

# Abstract

---

- Incorporate semantic knowledge for full parsing model
  - Collins style baseline model
  - Semantic knowledge Extraction
  - A class-based selection preference model to incorporate semantic knowledge
- Empirical analysis of semantic knowledge rules in ambiguity resolution
  - Nominal compounds disambiguation
  - Coordination disambiguation
  - N/V tagging disambiguation
  - Smoothing

# Introduction

---

- A lot of work on incorporating semantic knowledge for smaller tasks, like PP-attachment disambiguation.
- Difficulties of incorporating semantic knowledge into generative parsing model lie in:
  - Changing model structures to add additional features
  - Cleaning semantic data from external dictionaries
- One way to these problems: building a semantic sub-model to deal with semantic knowledge

# Collins style baseline model

---

- Similar to Collins Model I:  
decomposing lexicalized rules into  
smaller objects
  - Generating heads
  - Generating left/right Modifiers of head
    - Generating syntactic label and  
corresponding head tag
    - Generating new head words of modifiers

# Mixed new semantic sub-model with baseline model

---

- Using two sub-models to generate modifier head word
  - $P_{\text{mix}}(\text{mod headword}) = \lambda * P_{\text{org}} + (1 - \lambda) * P_{\text{sem}}$
  - Using EM algorithm to train parameter  $\lambda$

# Semantic dictionaries

---

## □ HowNet

- covers 67,440 words defined by 2112 different sememes
- With fine-grained ontology
- <http://www.keenage.com/>

## □ CiLin

- represents 77,343 words in a dendrogram
- Just a synset

# Extracting semantic categories

---

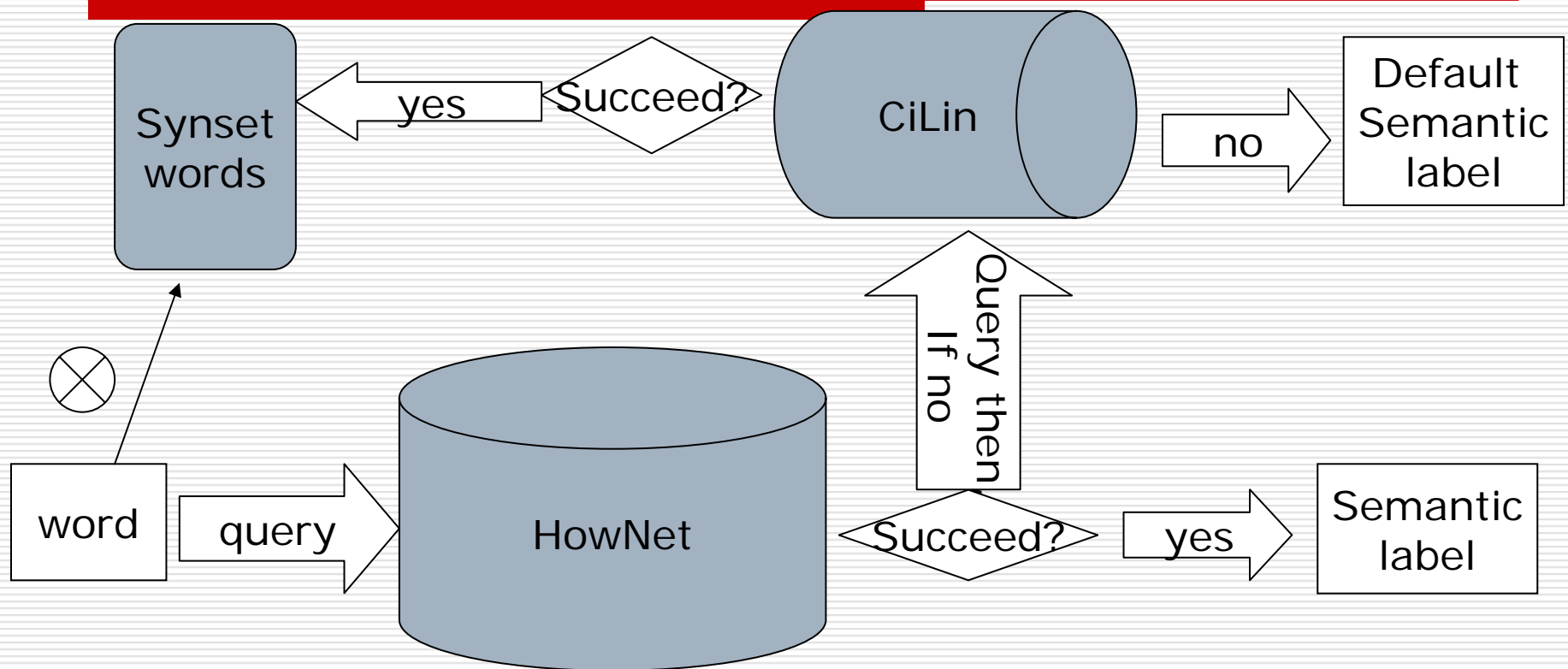
- From HowNet
  - Using the first definitions as semantic categories at the first level HN1
  - Through Hypernym ladders, extracting semantic categories at level HN2 and HN3 by replacing categories at lower level with their hypernyms at higher level.
- From CiLin
  - Organized words with three levels in total
  - 12 categories at first level, 97 categories at second level, 1400 categories at third level
- From heuristic rules
  - Numbers and time expressions
  - Just using suffixes and feature characters

# Word Coverage and semantic labels

---

	Data	HN1	HN2	HN3	CL1	CL2	CL3
words in train	9522	6040			6469		
words in test	1824	1538			1581		
words in both	1412	1293			1310		
classes in train	-	1054	381	118	12	92	1033
classes in test	-	520	251	93	12	79	569
classes in both	-	504	248	93	12	79	552

# Combination of HowNet, CiLin and heuristic rules



# Word Coverage and semantic labels for combination

---

	Data	HN2+HR	HN2+CL1+HR	HN2+CL2+HR	HN2+CL3+HR
words in train	9522	7347	7911	7911	7911
words in test	1824	1608	1672	1672	1672
words in both	1412	1343	1372	1372	1372
classes in train	-	381	381	381	381
classes in test	-	251	251	251	251
classes in both	-	248	248	248	248

# Sub-model for incorporating semantic knowledge

---

## □ Bilexical-class sub-model

- $P(\text{seml}(\text{modheadword}) | \text{seml}(\text{headword}))$
- Too coarse-grained to capture semantic preferences: different words in the same semantic category impose different selection restrictions on their modifier words

## □ Bilexical-word dependency

- $P(\text{modheadword} | \text{headword})$
- Too fined grained and result in sparseness

# Class-based selection preference model

---

- Model the selection restrictions imposed by semantic class of headword on modifier headword
- $P(\text{modheadword} | \text{seml}(\text{headword}))$
- Just fit to capture semantic dependencies

# Experimental setup

---

- CTB release 1.0
- For the EM algorithm
  - Articles 1-25 as held out data
  - Articles 26-270 as training data
- For the final test
  - Articles 1-270 as training data
  - Articles 271-300 as test data

# Results on HowNet and CiLin respectively

---

	Baseline	HN1	HN2	HN3	CL1	CL2	CL3
F1(%)	78.5	78.6	79.1	78.9	77.5	78.7	78.8

# Results on the combination of HowNet and CiLin

---

	Baseline	HN2+CL1+HR	HN2+CL2+HR	HN2+CL3+HR
F1(%)	78.5	79.2	79.4	79.3

# Significance test

---

- Computationally-intensive randomization test
- A 1,048,576 trial approximate randomization is performed twice to test the improvement of F1 score
- The significance level is  $(43376 + 1) / (1048576 + 1) = 0.041$

# Empirical study of results

---

- Three main types of errors
  - NP-NP modification
  - Coordination
  - N/V tagging

# Stats about three errors

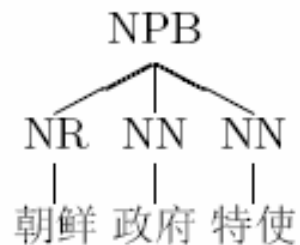
---

Type	Count	Percent(%)
Nominal Compound <i>AR</i>	19	38
Coordination <i>AR</i>	9	18
$N \diamond V$ <i>AR</i> in $N \diamond V + noun$	6	12
Other <i>AR</i>	16	32

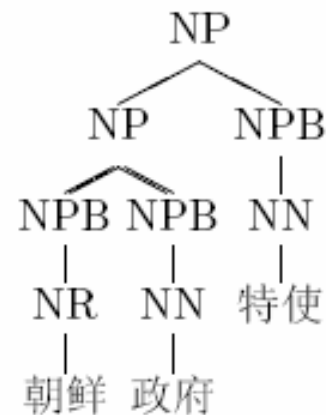
# Nominal compounds disambiguation

---

a.



b.



# Coordination disambiguation

---

- The latter VP is incorrectly parsed as IP due to pro-drop
- Selection preference model penalize this parse by the meaning similarity of coordinated structures.

# N/V tagging disambiguation

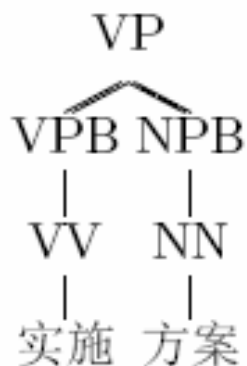
---

- Always tagged incorrectly in the pattern N/V + noun
  - If N/V is verb, the pattern is parsed as predicate-object structure
  - Else N/V is noun, parsed as modifier-noun structure
- Semantic preferences are helpful for parsing the pattern

# N/V tagging disambiguation

---

a.



b.



# Smoothing

---

- alleviate the fundamental sparseness of the lexical dependency information available in the treebank
- Many other structures disambiguation because of semantic knowledge's role of generalization

# Conclusions and future work

---

- Semantic knowledge is helpful for parsing
  - NP-NP modification disambiguation
  - Coordination disambiguation
  - N/V tagging disambiguation
  - Smoothing disambiguation
- Using semantic knowledge in discriminative parsing models