Developing Probabilistic Models for Identifying Semantic Patterns in Texts

Minhua Huang and Robert M. Haralick

Computer Science, Graduate Center
The City University of New York
New York, NY 10016

September 18, 2011

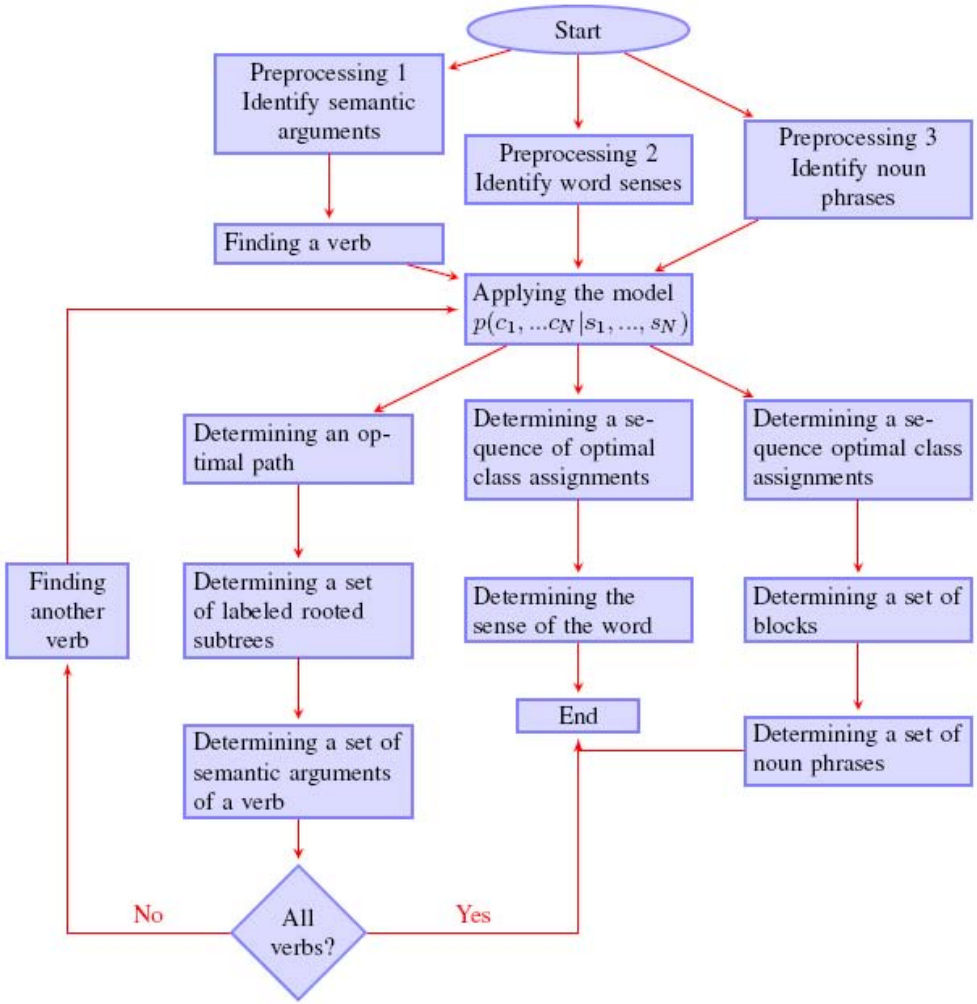
Introduction

Three text patterns capturing semantics of a sentence

- Semantic arguments of a verb
  - Semantic arguments of a verb can be used to answer the questions of who, what, when, where, and why.
- The meaning of a word
  - The sense of a polysemous word can be used to understand the meaning of the word.
- Noun phrases
  - Noun phrases of a sentence combining with verbs can be used to find the abstraction of the sentence.

The Algorithm

- The key of the algorithm is a probabilistic graphical model.



The Probabilistic Graphical Model

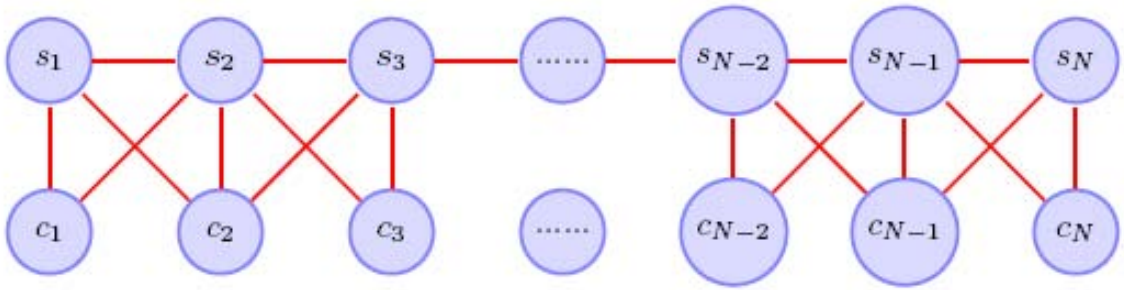


Figure: The conditional independence graph defining our graphical model.

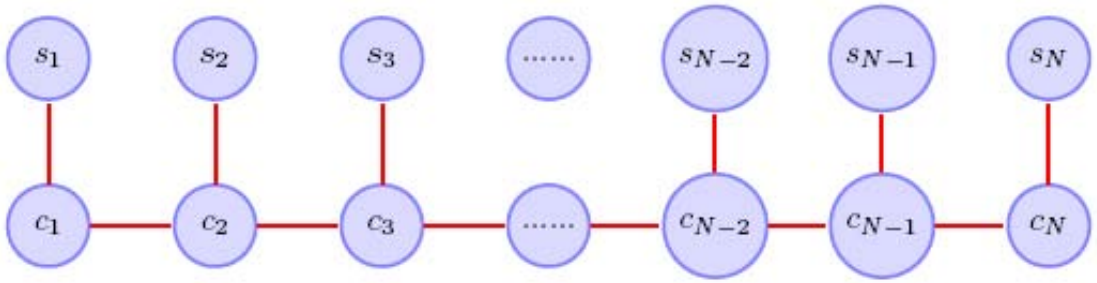
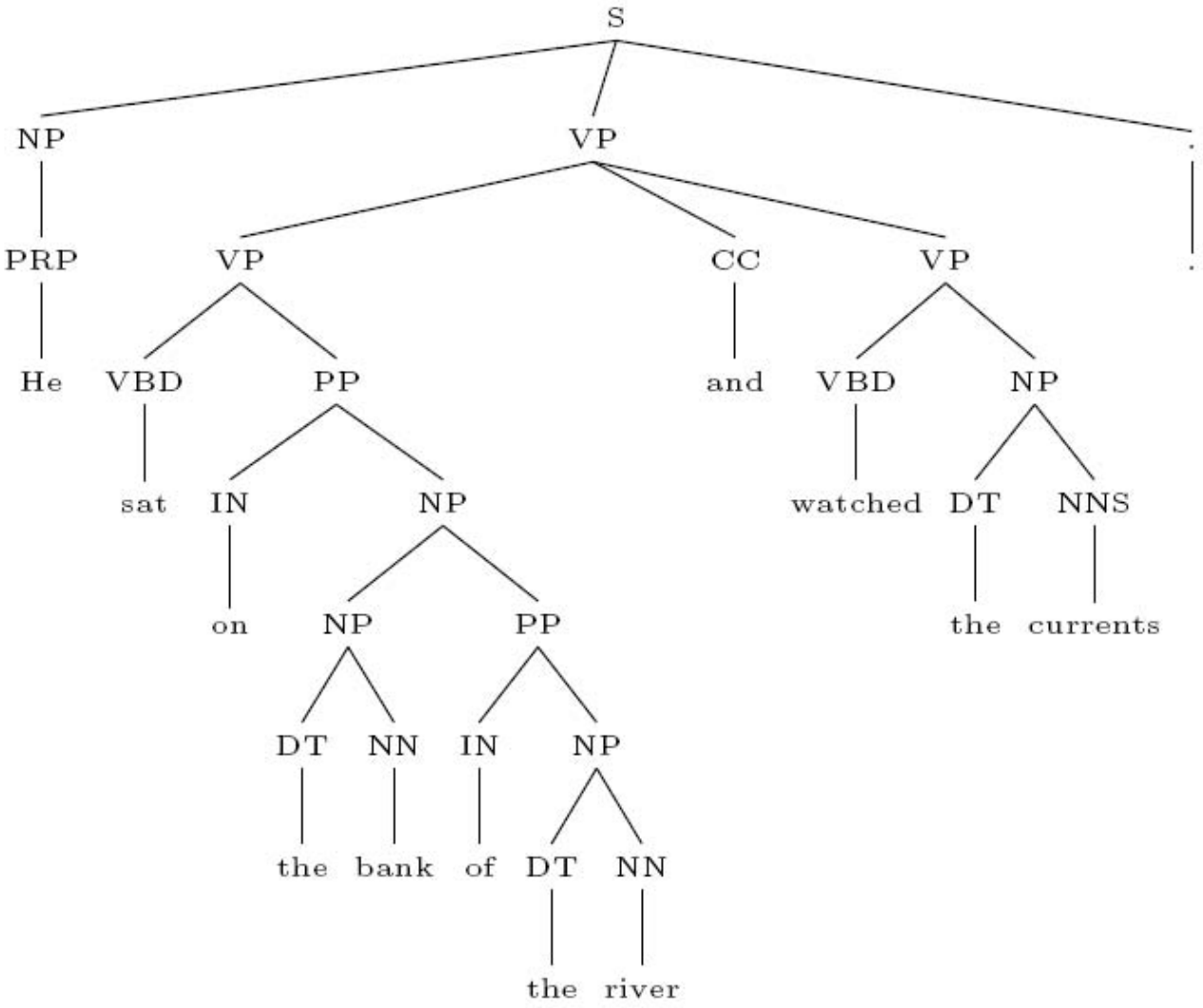


Figure: The usual conditional independence graph for Markov dependencies among the classes.

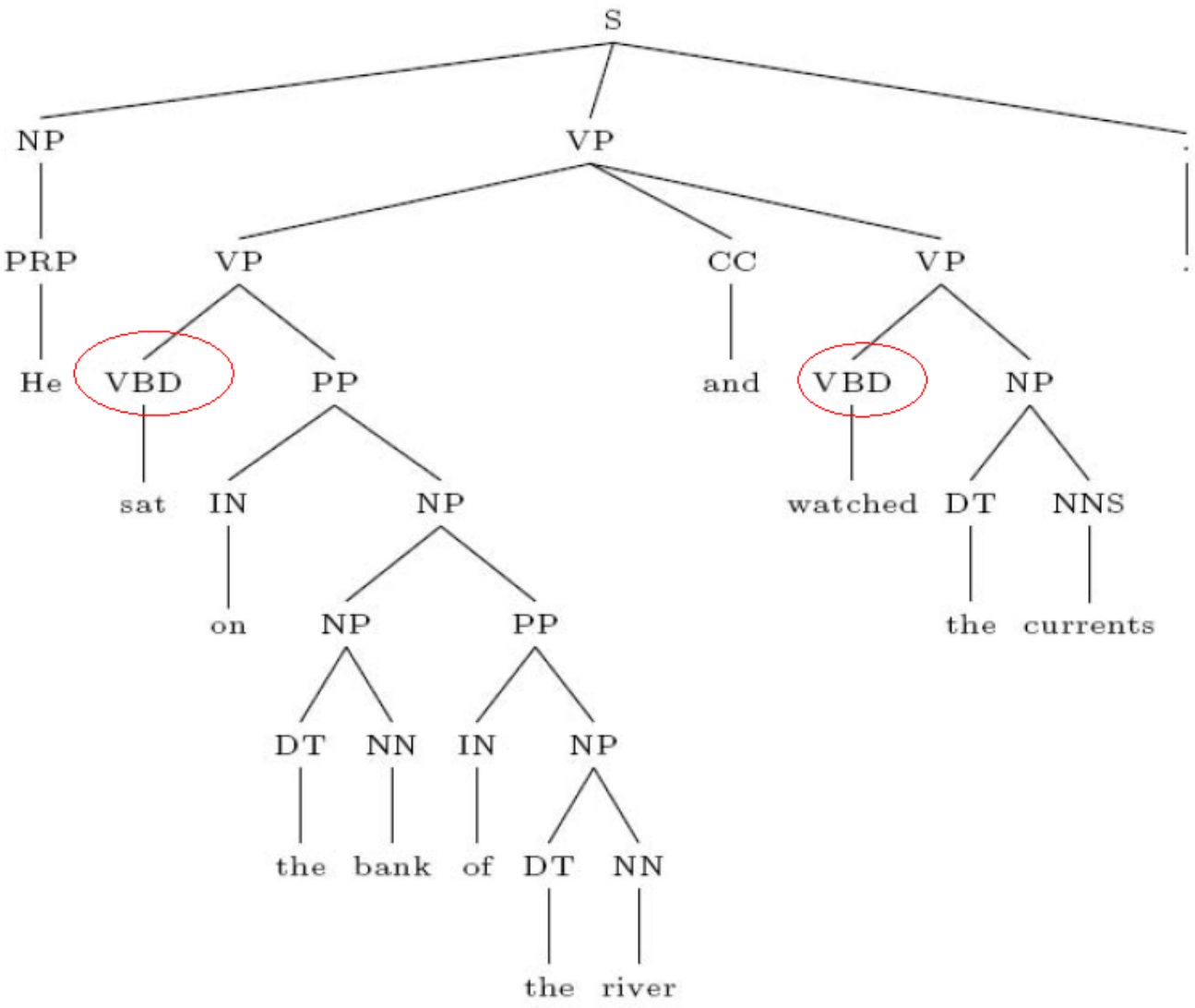
An example of identifying semantic arguments of a verb

He sat on the bank of the river and watched the currents.



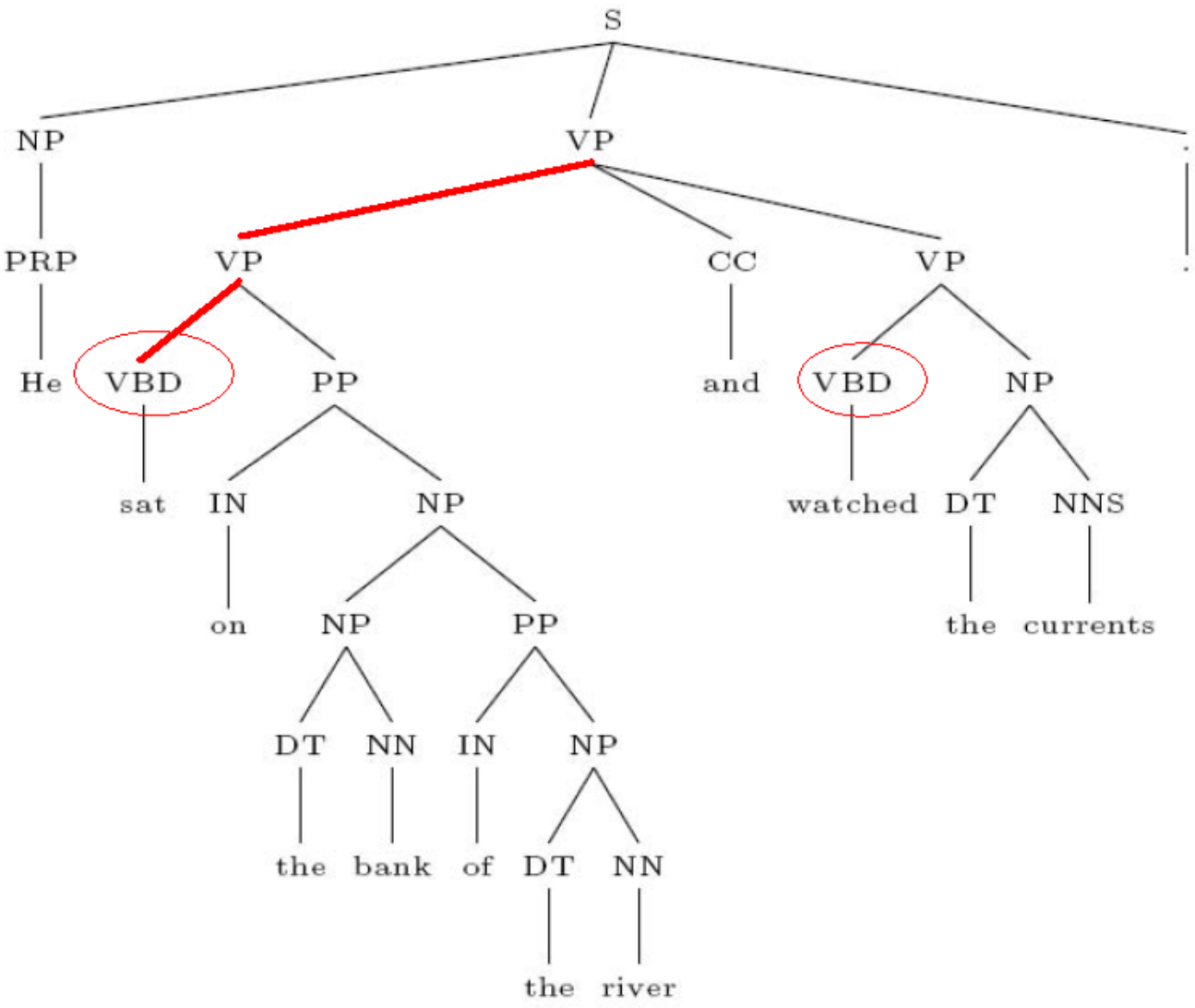
An example of identifying semantic arguments of a verb

He sat on the bank of the river and watched the currents.



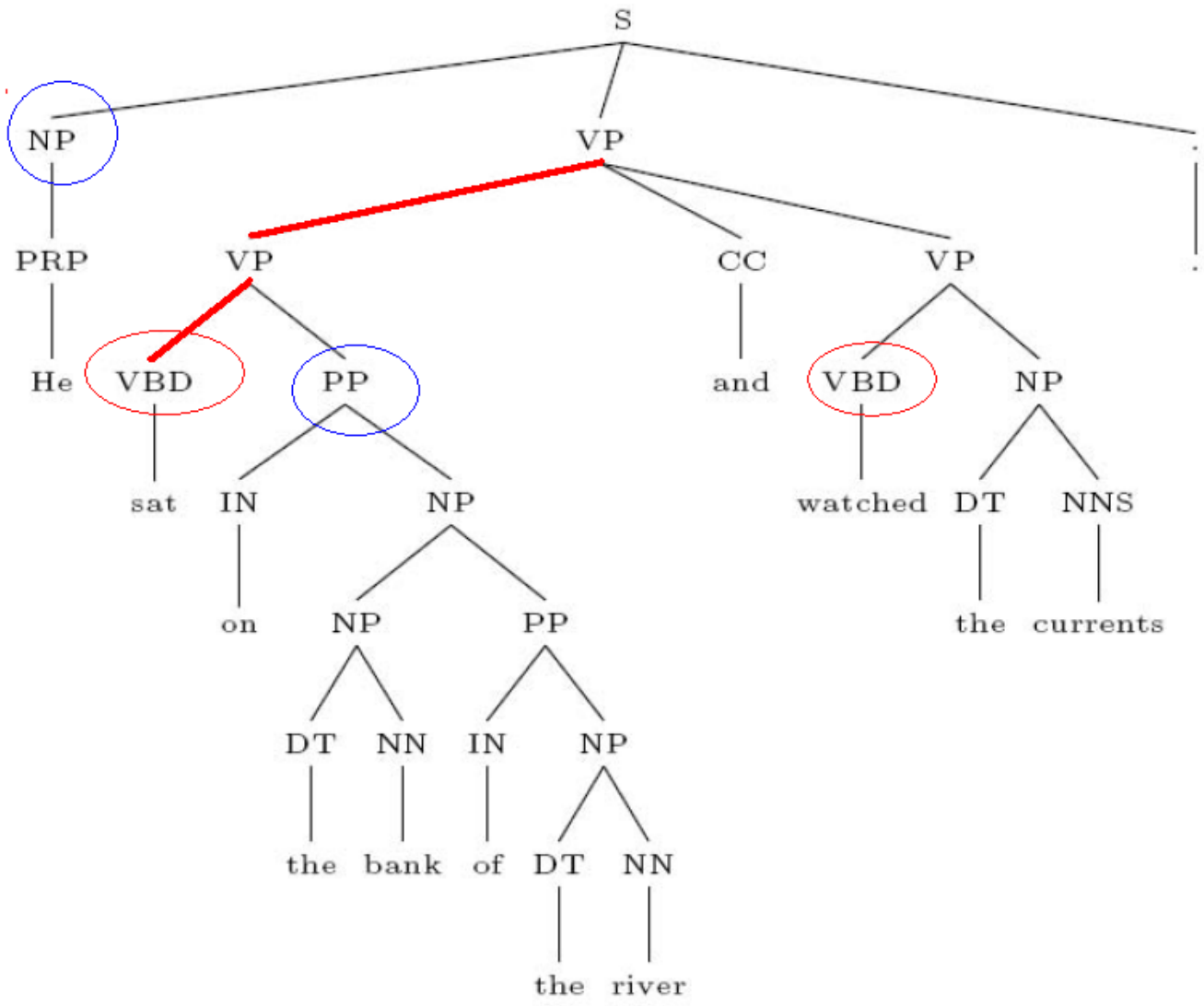
An example of identifying semantic arguments of a verb

He sat on the bank of the river and watched the currents.



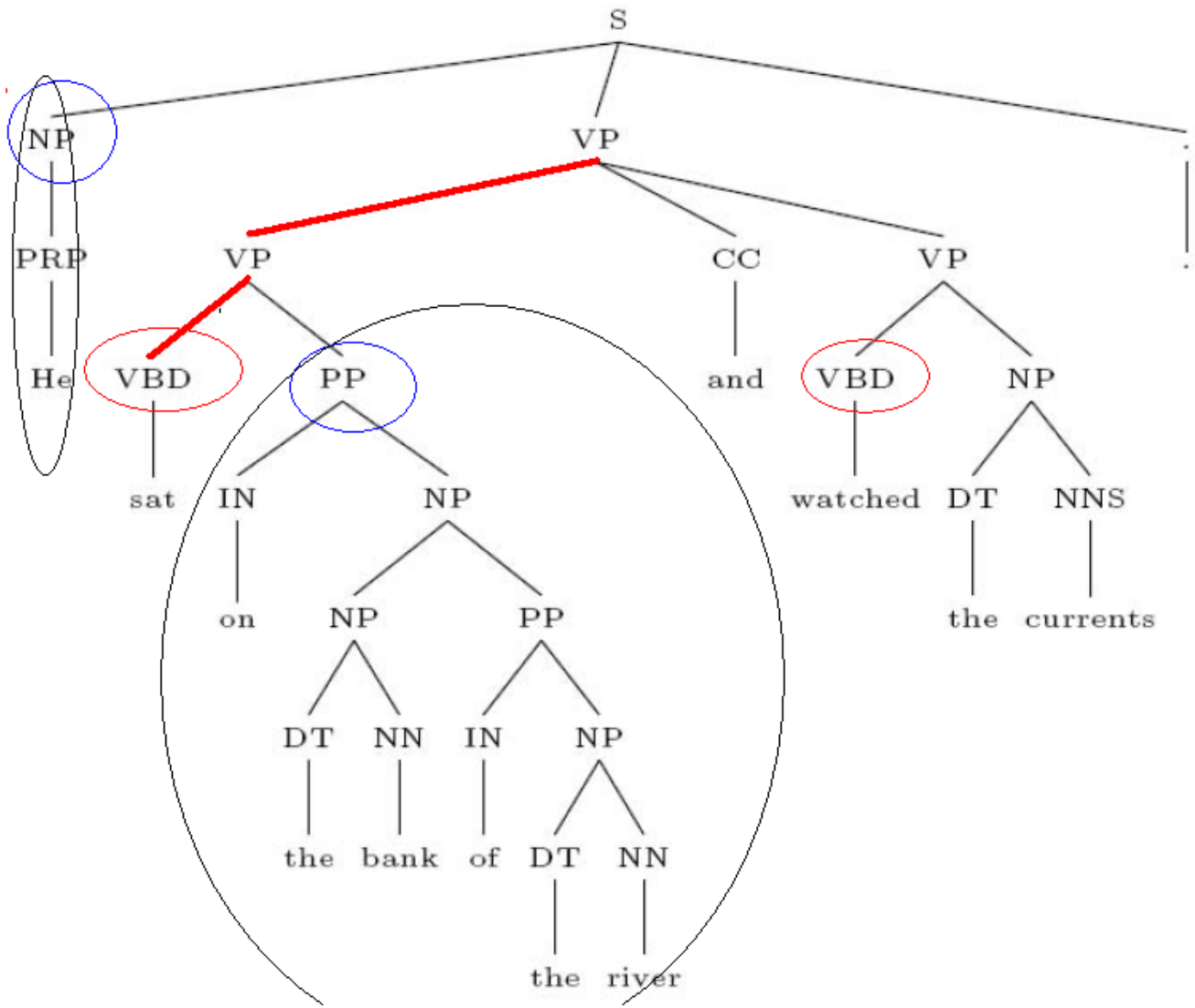
An example of identifying semantic arguments of a verb

He sat on the bank of the river and watched the currents.



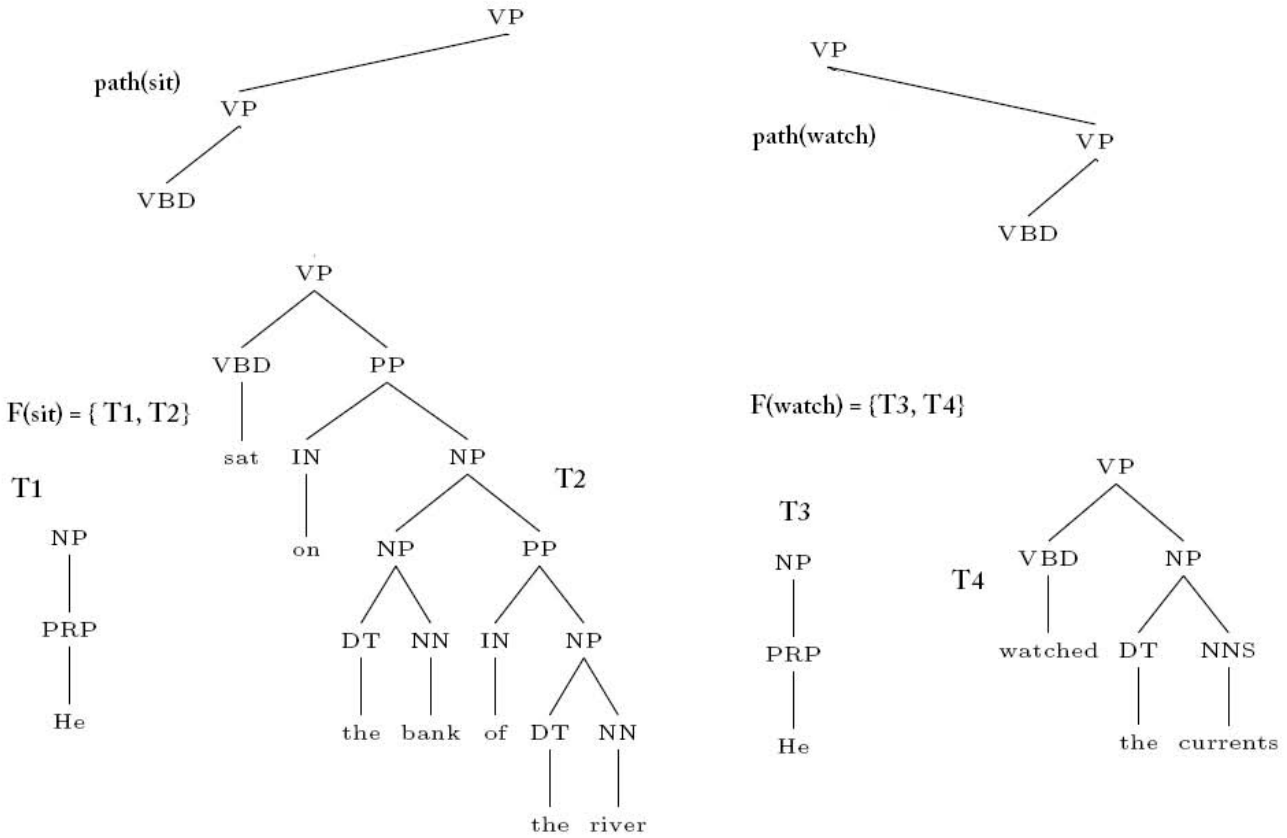
An example of identifying semantic arguments of a verb

He sat on the bank of the river and watched the currents.



An example of identifying semantic arguments of a verb

He sat on the bank of the river and watched the currents.



- Semantic arguments:
  - sit: *he; on the bank of the river*
  - watch: *he; the currents*

Empirical Results

- Results for identifying semantic arguments of a verb in a sentence on *WSJ* data from Penn Treebank and PropBank
- About 600 verbs associating with about 2000 semantic arguments
- 10 – *fold* cross validation technique

Files 20, 37, 49, 89	Precision %	Recall %	F-Measure %
Average	92.335	94.1675	93.2512
Standard Deviation	0.6195	0.5174	0.4605

Empirical Results

- Results for identifying the meaning of a word in a sentence on *line* data
- Results are better than those published by other researchers
- 10 – *fold* cross validation technique

	Accuracy 3 senses %	Accuracy 6 senses %	# of Context words in Training Set 6 senses k	Base Line 6 senses %
LSA [11]	75			
Bayesian [9]	76	71	8.9	16.67
Context Vector [9]	73	72	8.9	16.67
Neural Network [9]	79	76	8.9	16.67
This Method	85.25	81.12	2.45	19.09

Empirical Results

- Results for identifying noun phrases in a sentence on *CoNLL* – 2000 data
- Results are better than *The-Context-Independent-Bayes* model
- Results are better than those published by other researchers

Method	Recall %	Precision %	F-measure %
Role Based Learning [12]	92.03	91.05	91.54
HMM [1]	93.52	93.43	93.48
Naive Bayes			93.69
MEMM [13]	–	–	93.70
Voted perceptrons [14]	93.29	94.19	93.74
CRF [13]	–	–	94.38
SVM [15]	94.38	94.52	94.45
our method [16]	95.31	96.36	95.74