

# A Few Thoughts and Attempts on the Future of SMT

Deyi Xiong

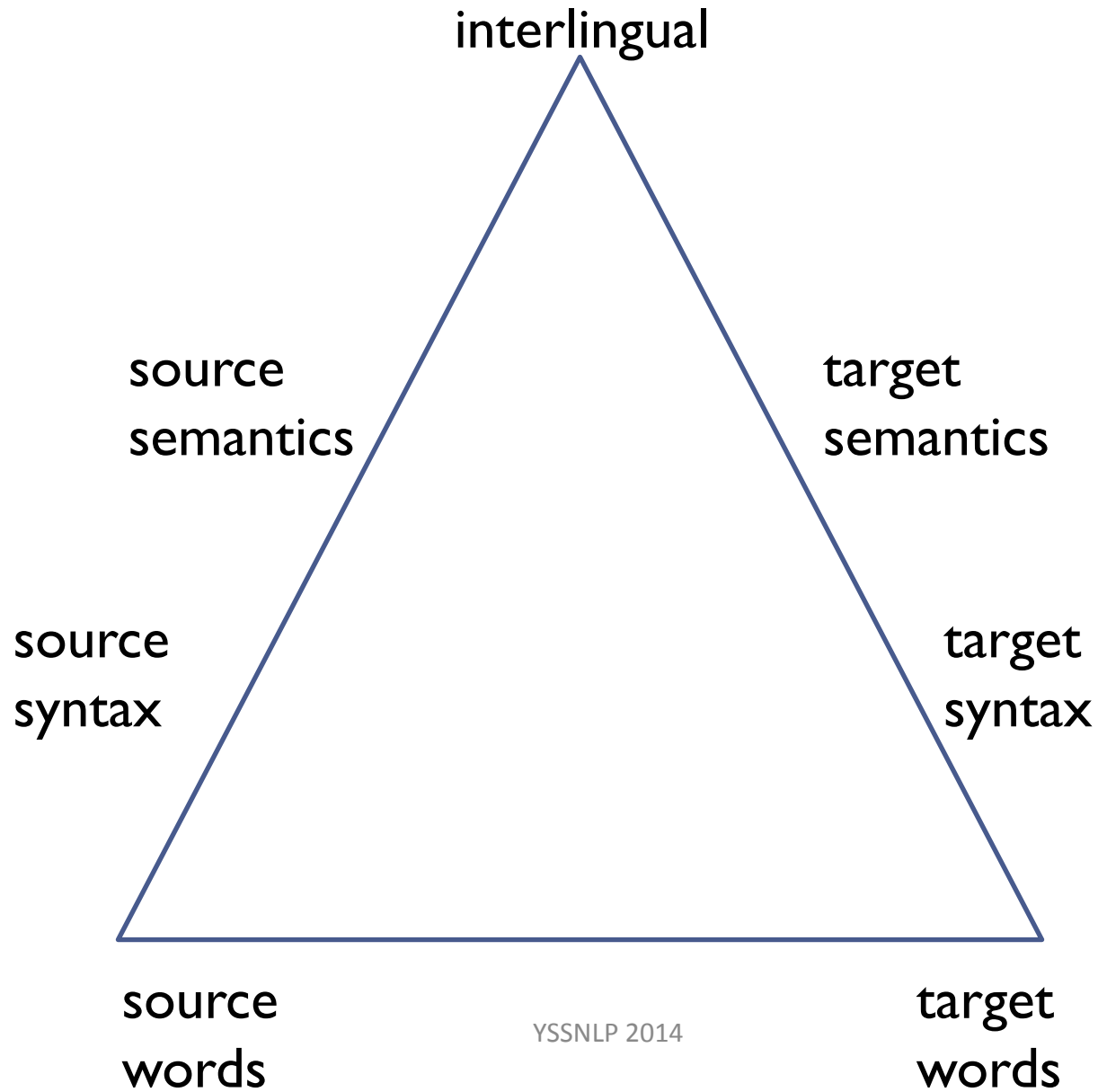
Soochow University, Suzhou, China



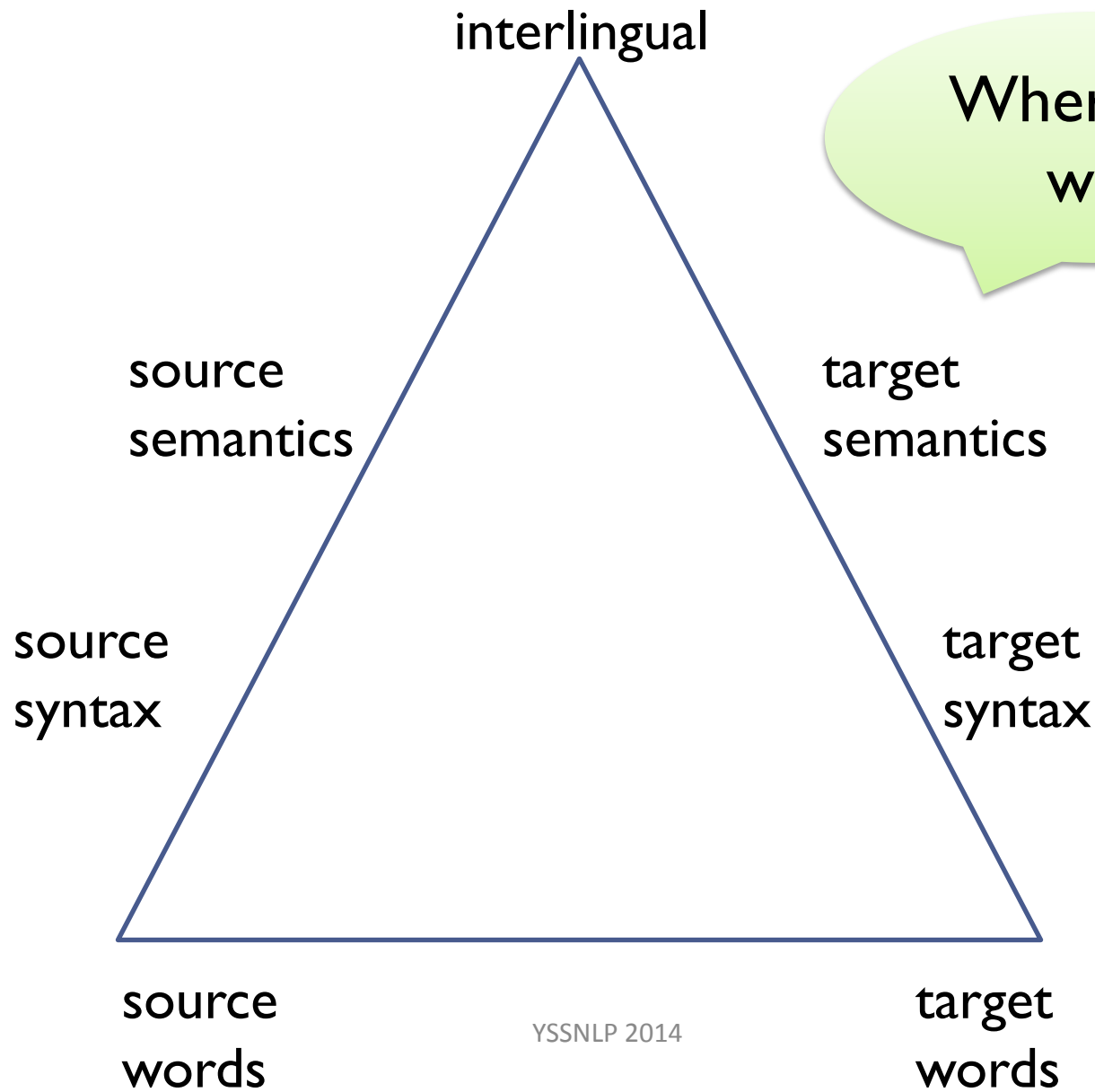
# Tower of Humankind



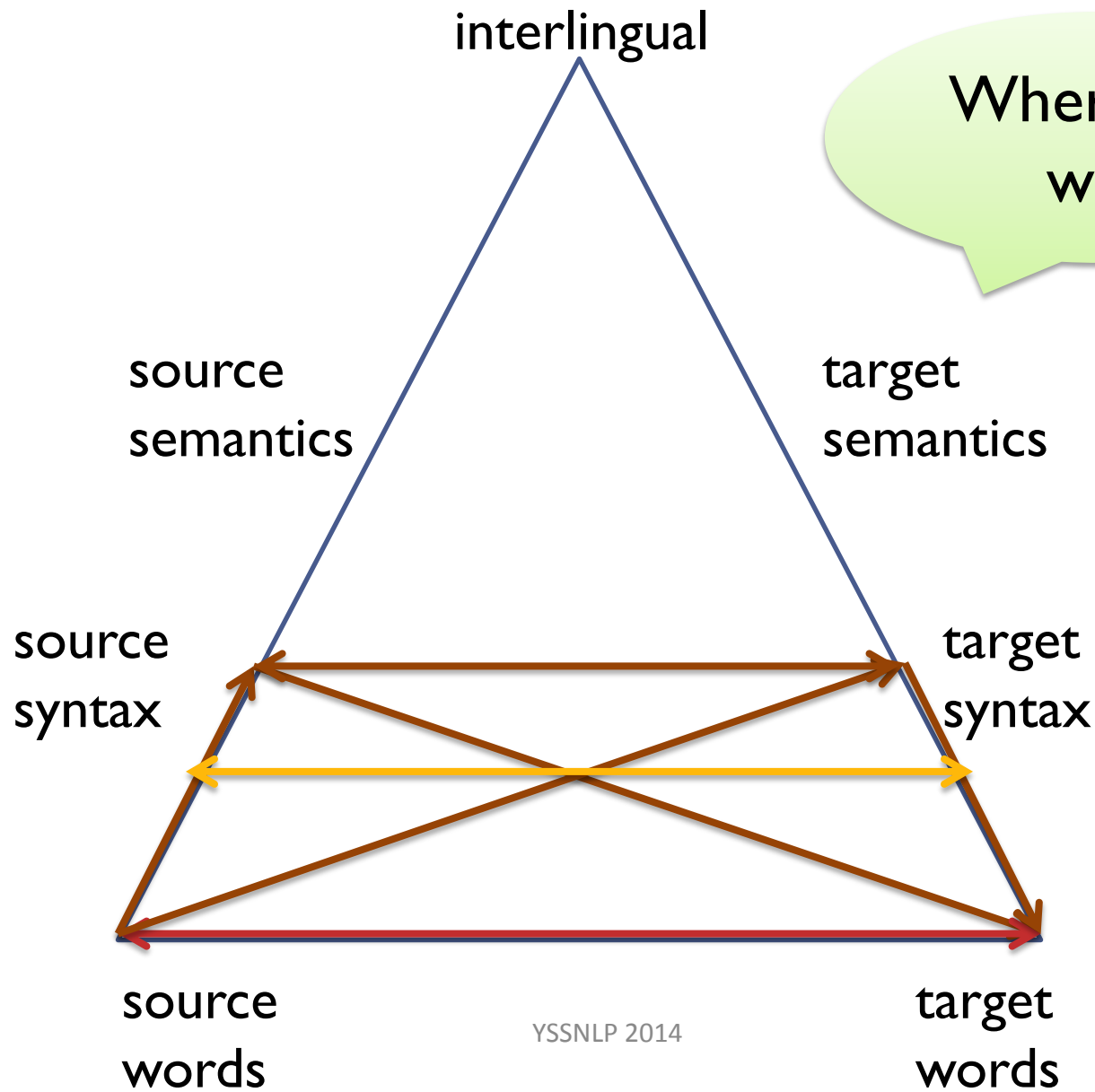
# Tower of Machine Translation



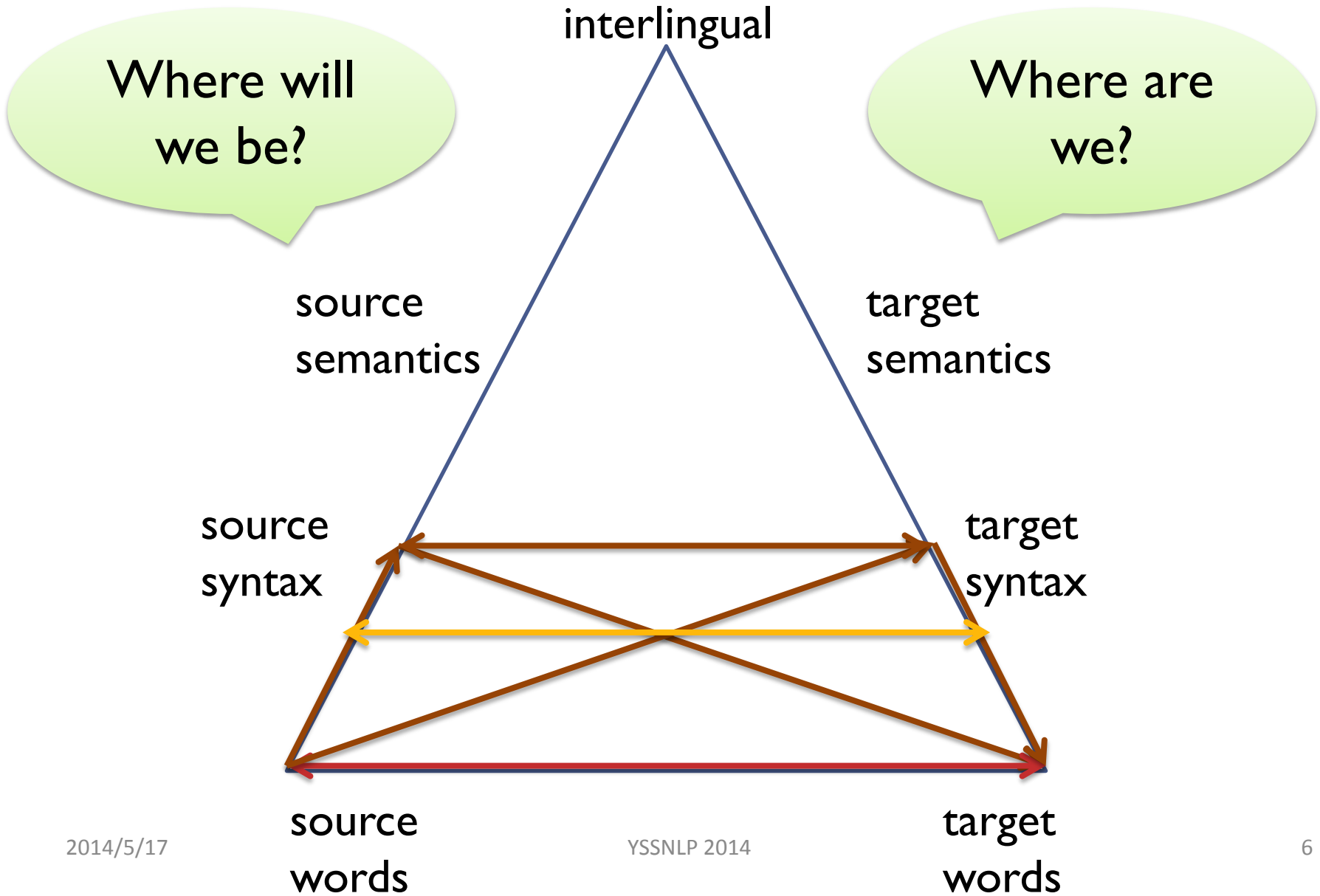
# Tower of Machine Translation



# Tower of Machine Translation



# Tower of Machine Translation



# SMT: Future Directions

- ▶ Representation
- ▶ Unsupervised learning
- ▶ Semantics, discourse and SMT

# Representation

- ▶ Knowledge representation
  - # one of central concepts in AI
  
- ▶ Translation representation
  - # word pairs → word-based SMT
  - # bilingual phrases → phrase-based SMT
  - # ITG rules → ITG-based SMT
  - # various tree-based rules → tree-based SMT



# Representation

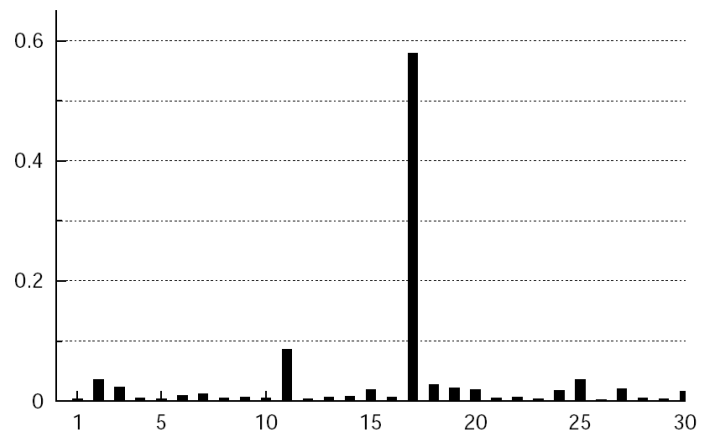
## ▶ What's next?

- ⌘ distributional representation
- ⌘ semantics-based whole-sentence representation
- ⌘ other new representations ...

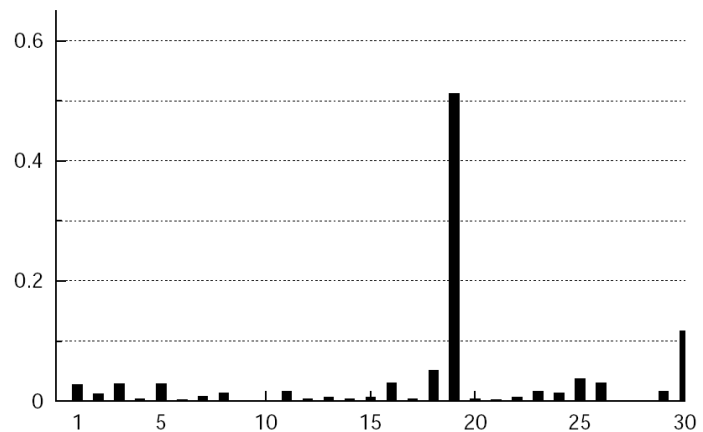
## ▶ Tradeoff

- ⌘ computational tractability
- ⌘ expressive capacity

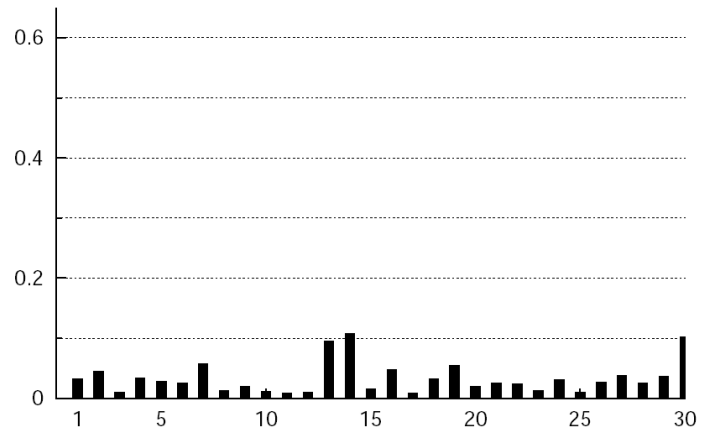
# Rule representations with topic distributions: Xiao et al., 2012



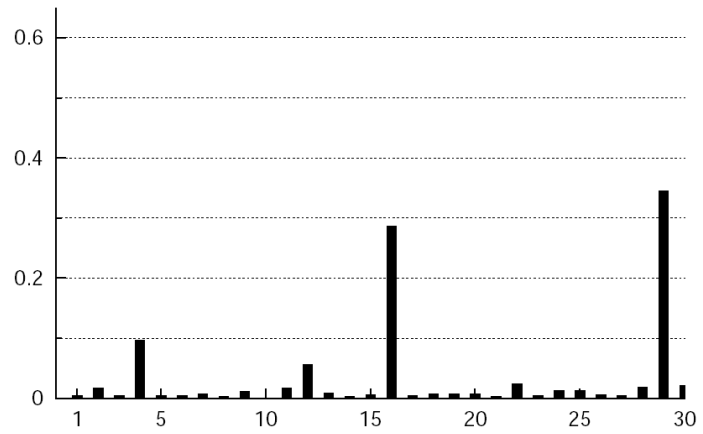
(a) 作战能力  $\Rightarrow$  operational capability



(b) 给予  $X_1 \Rightarrow$  grants  $X_1$

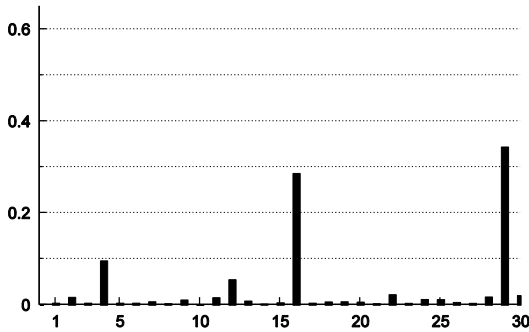


(c) 给予  $X_1 \Rightarrow$  give  $X_1$

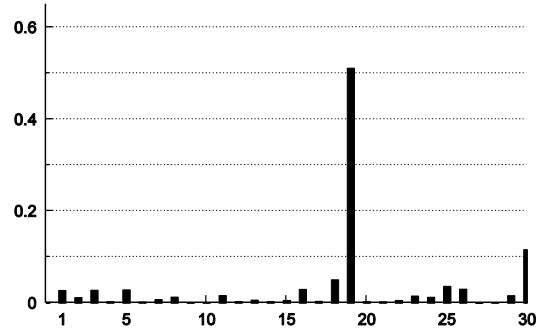


(d)  $X_1$  举行会谈  $X_2 \Rightarrow$  held talks  $X_1 X_2$

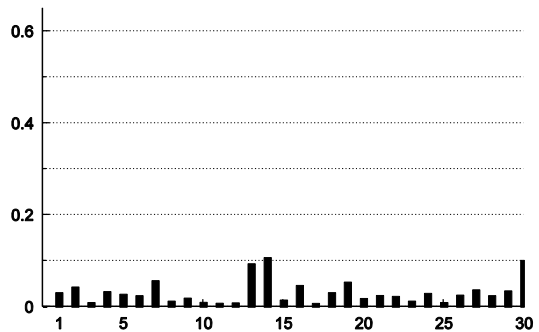
# Topic Similarity Model



Source Document

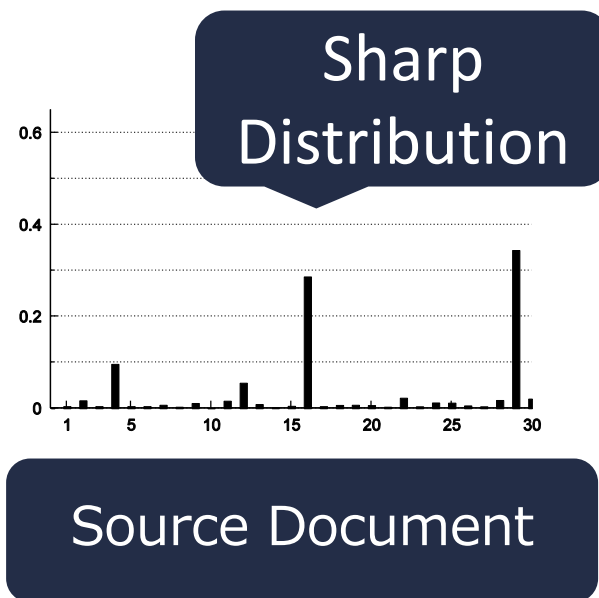
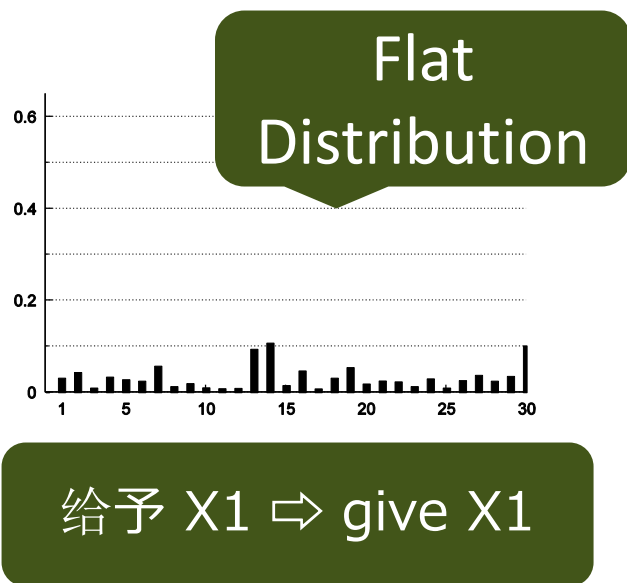


给予 X1 ⇔ grants X1



给予 X1 ⇔ give X1

# Topic Similarity Model: Bias

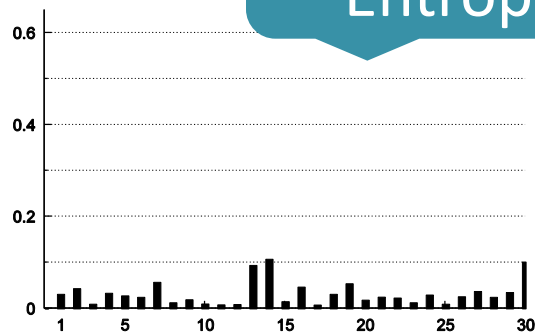


- ▶ Flat distribution is dissimilar to sharp distribution
- ▶ Topic similarity model: **punish** rules with even distributions
  - Topic-insensitive rule

# Topic Sensitivity Model

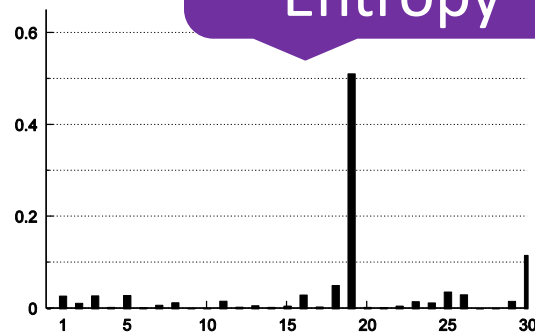
- ▶ Calculate entropy of rule-topic distributions of rules
- ▶ Compensate the bias of the topic similarity model

Large Entropy



给予 X1  $\Leftrightarrow$  give X1

Small Entropy



给予 X1  $\Leftrightarrow$  grants X1

# SMT: Future Directions

- ▶ Representation
- ▶ **Unsupervised learning**
- ▶ **Semantics, discourse and SMT**

# Unsupervised Learning

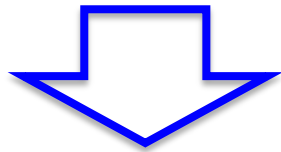
- ▶ EM for word alignment: big success
- ▶ EM for phrase/grammar induction: failure or moderate success
- ▶ Unsupervised discriminative learning: an exciting direction for grammar induction (Xiao et al., 2012; Xiao and Xiong 2013)

# Max-Margin Synchronous Grammar

## Induction: Xiao and Xiong 2013

奥巴马 与 内塔尼亚胡 举行 会谈

Obama hold a talk with Netanyahu



奥巴马  $X_1$   $\rightarrow$  Obama  $X_1$

与 内塔尼亚胡  $\rightarrow$  with Netanyahu

$X_1$  举行 会谈  $\rightarrow$  hold a talk  $X_1$

.....

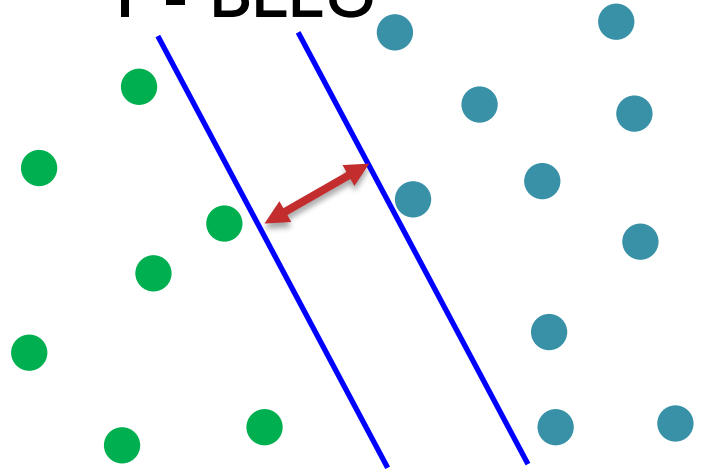




# Max-Margin SGI

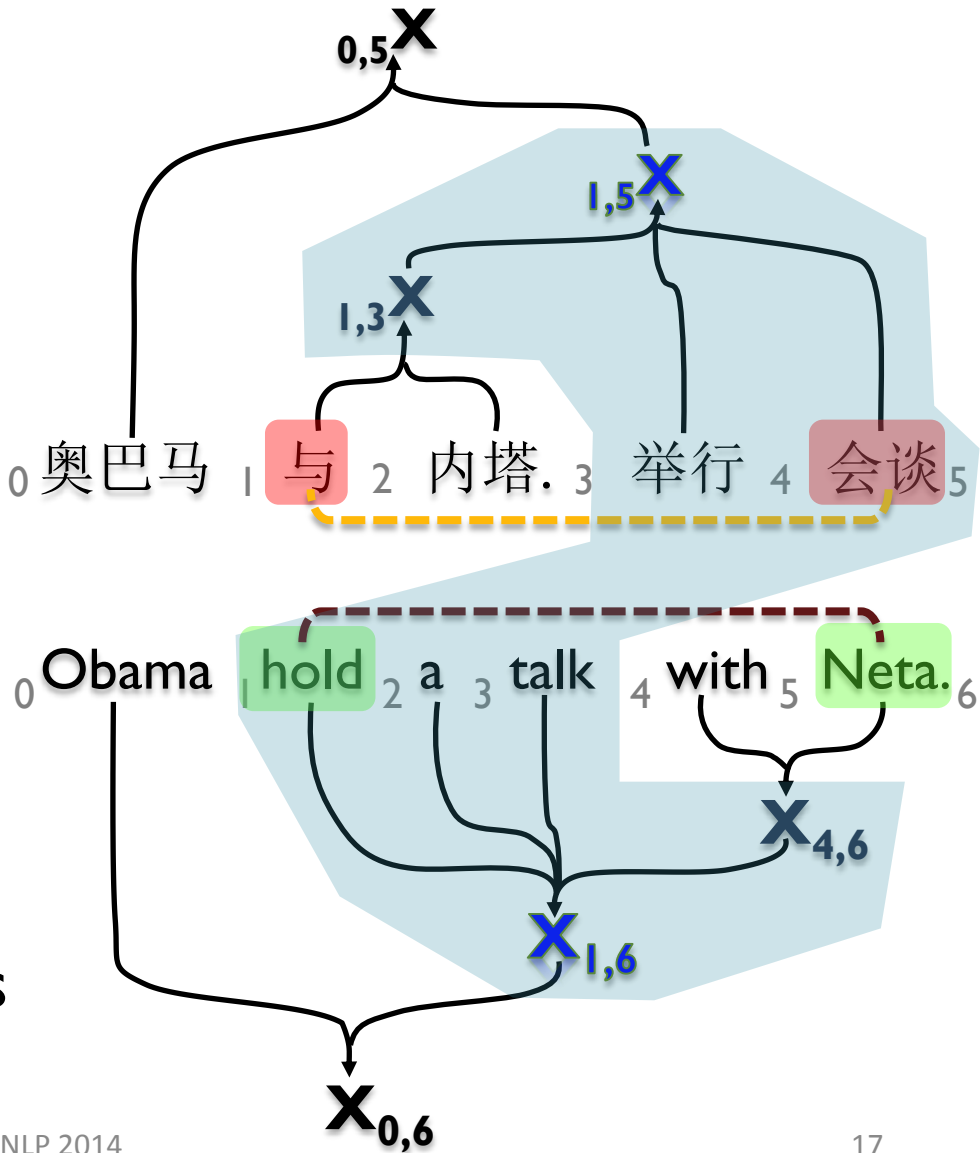
▶ Max-margin

I - BLEU



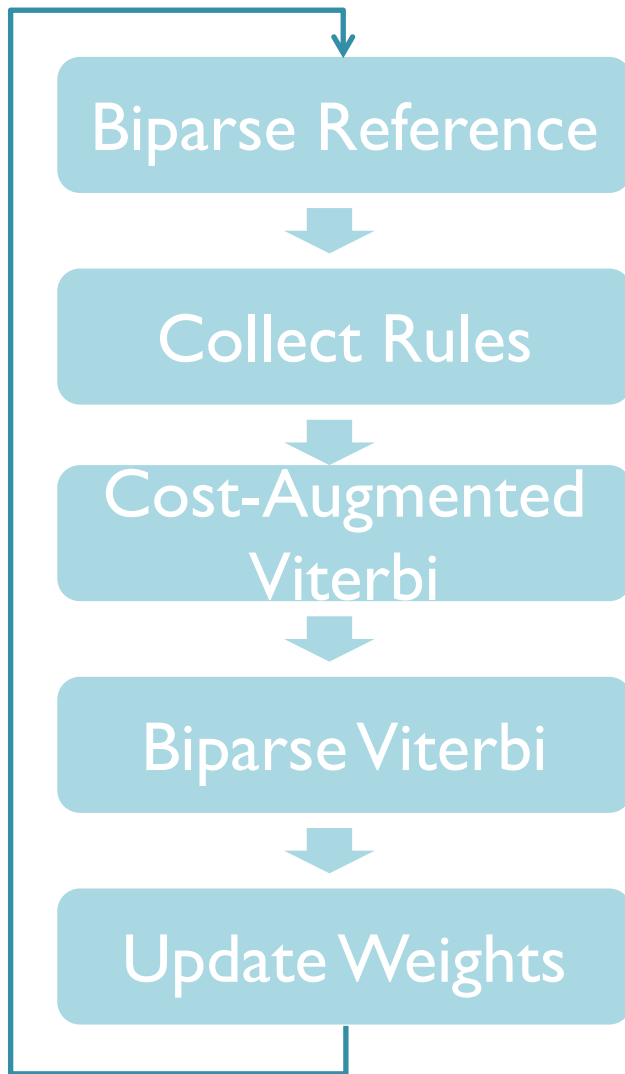
▶ Non-local features

# Target Parse Structures



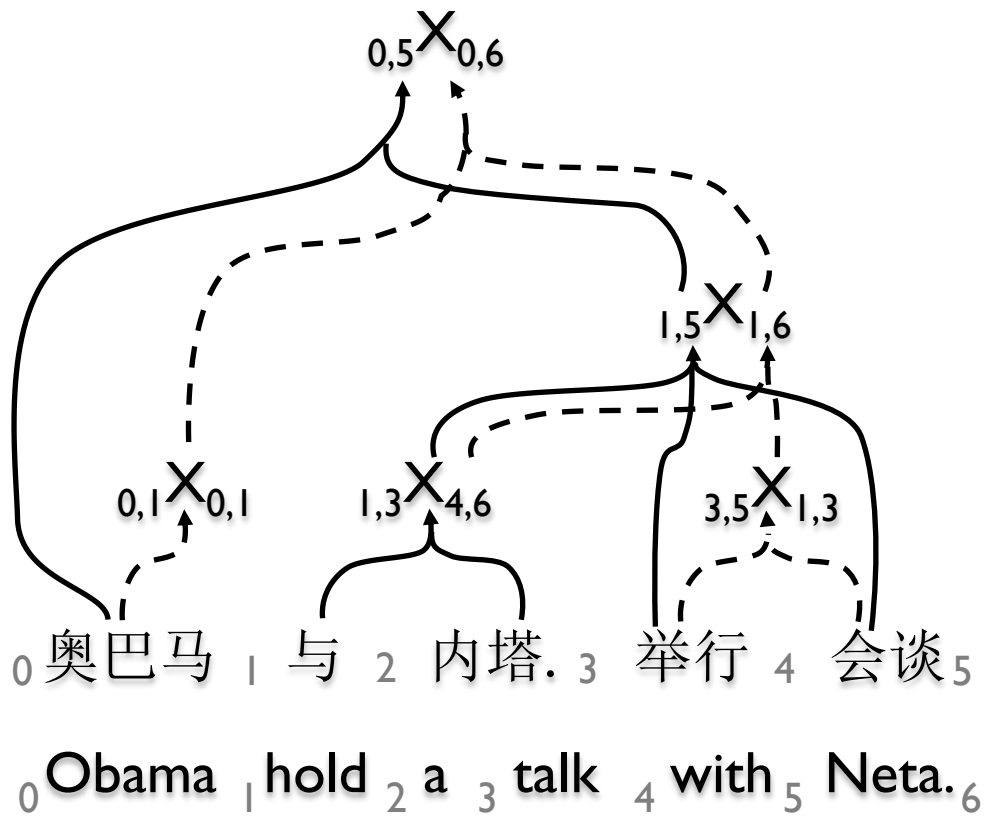
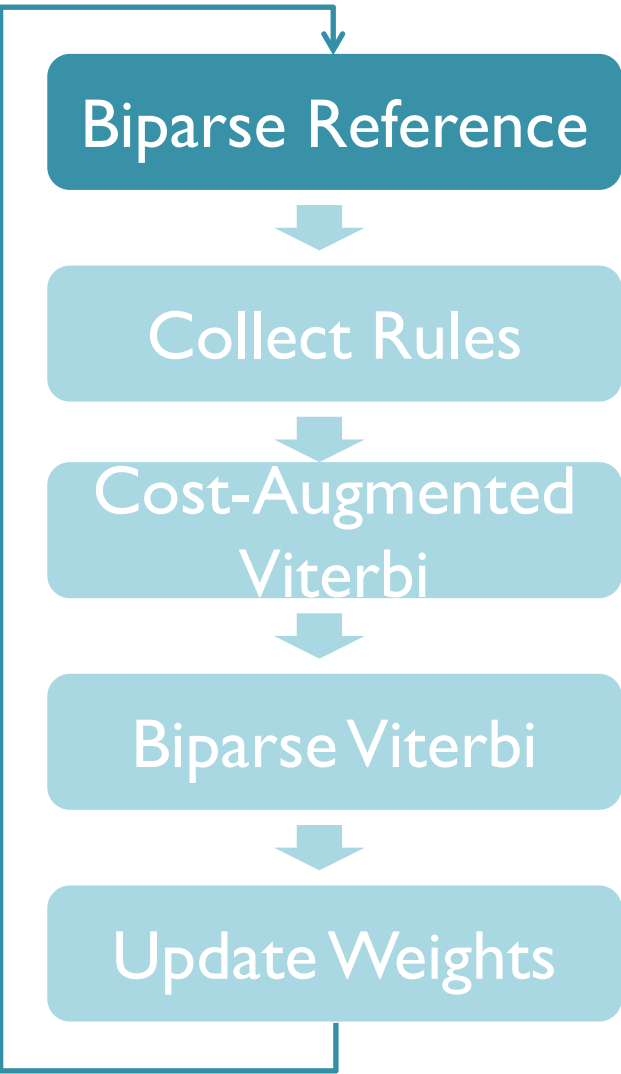
# Optimization

For each sent.



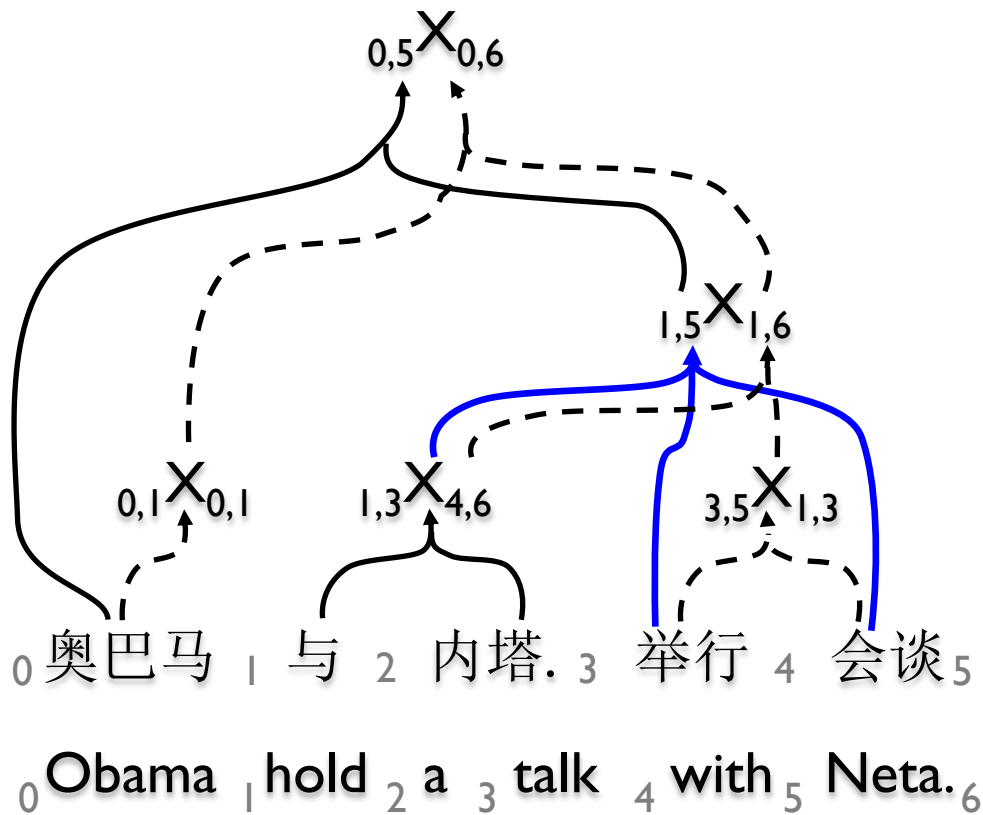
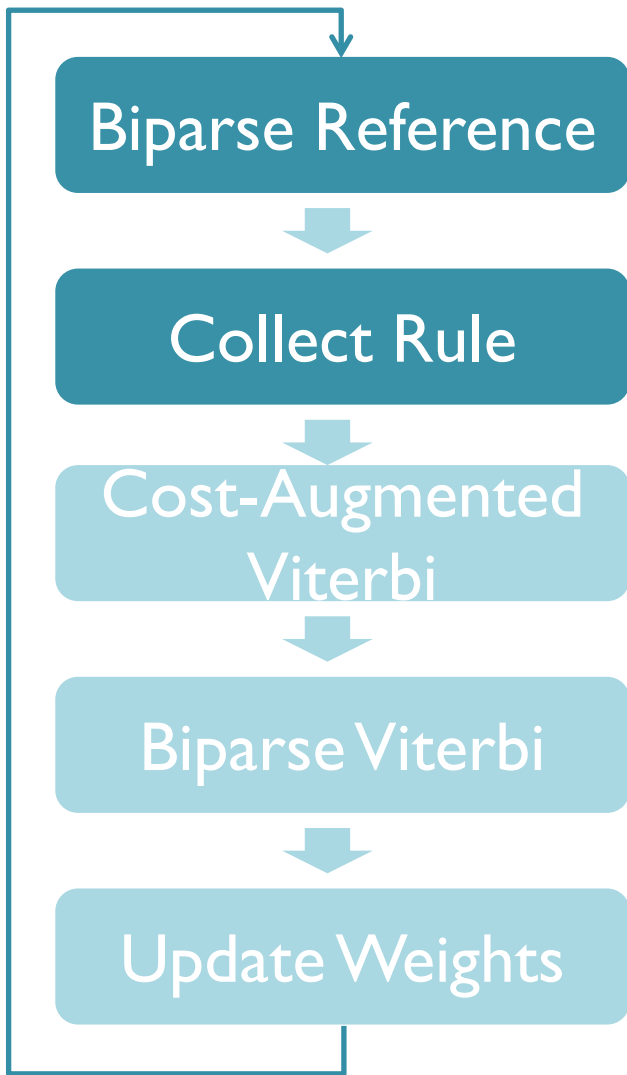
# Optimization

For each sent.



# Optimization

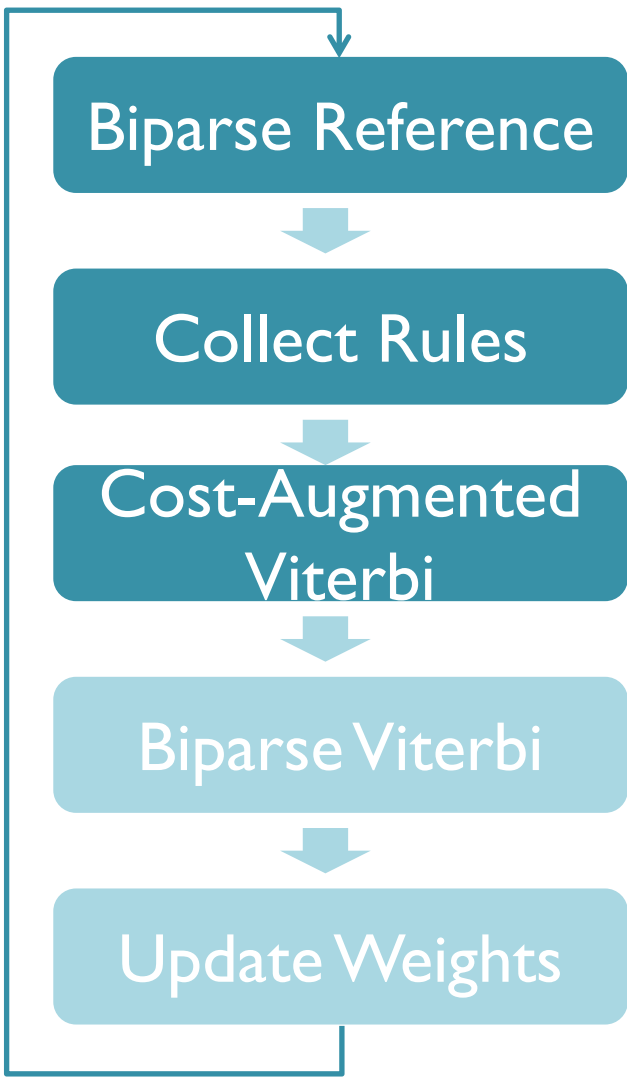
For each sent.



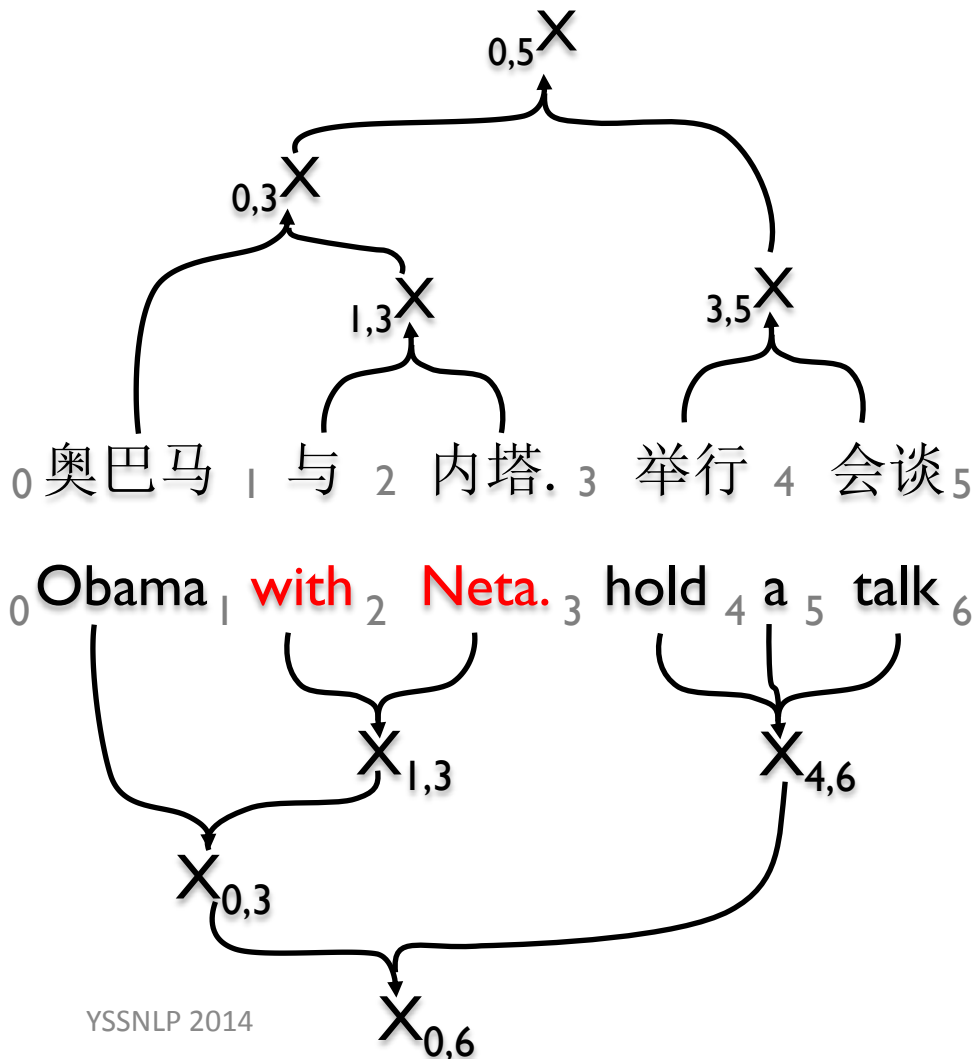
$X_1$  举行 会谈  $\rightarrow$  hold a talk  $X_1$

# Optimization

For each sent.

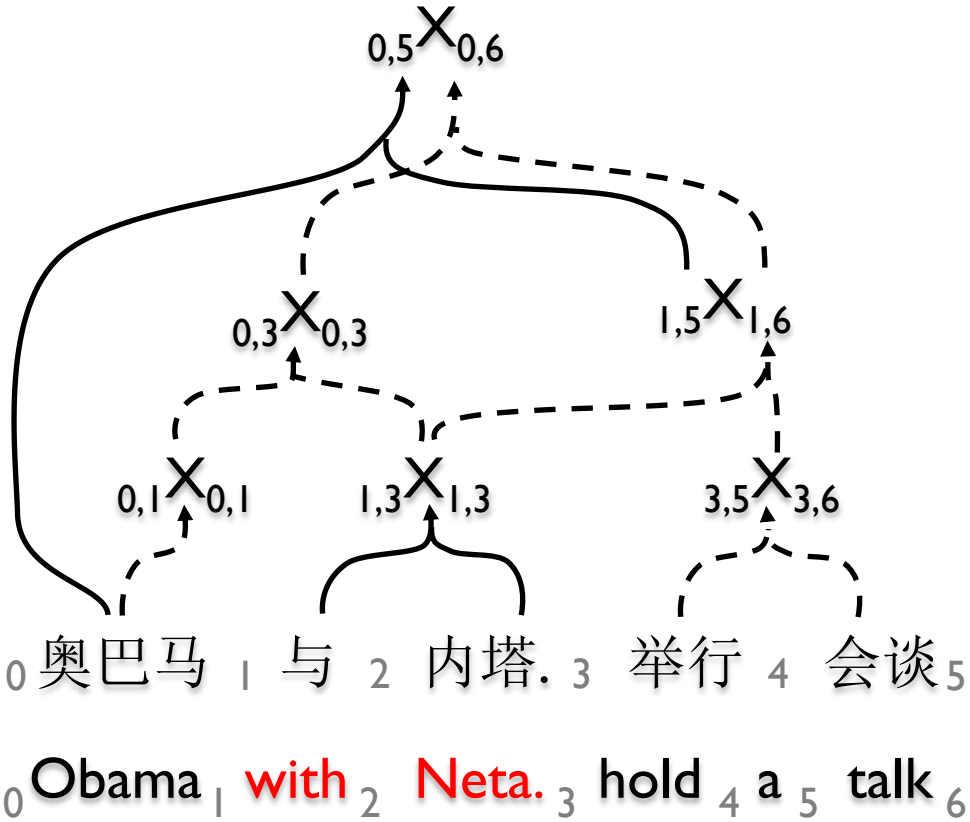
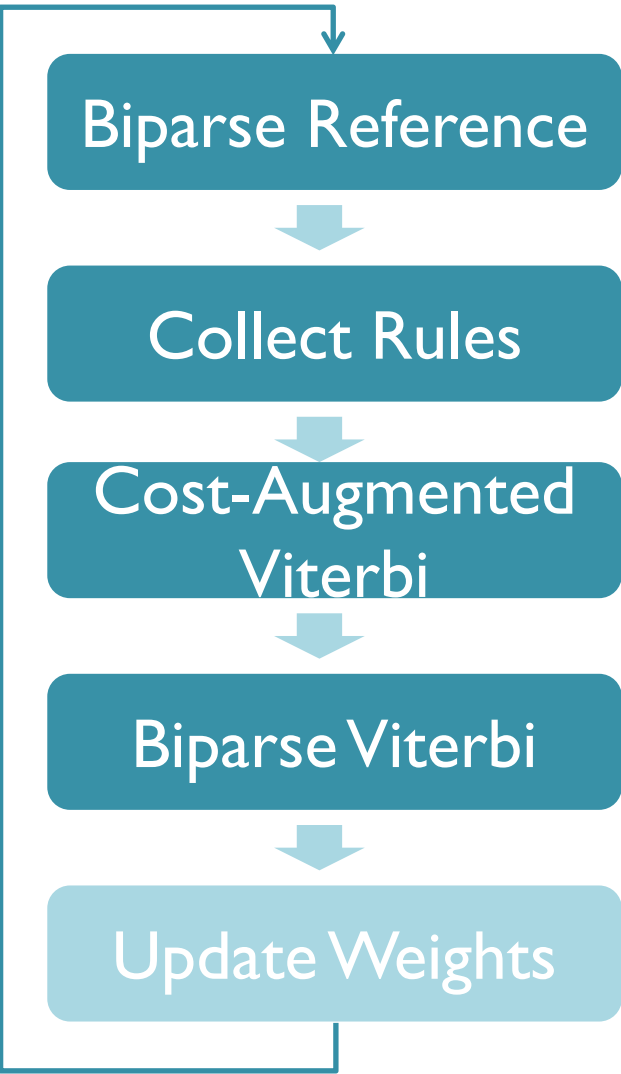


- $f(\mathbf{s}^{(i)}, \mathbf{t}) - \text{BLUE-4}(\mathbf{t}^{(i)}, \mathbf{t})$



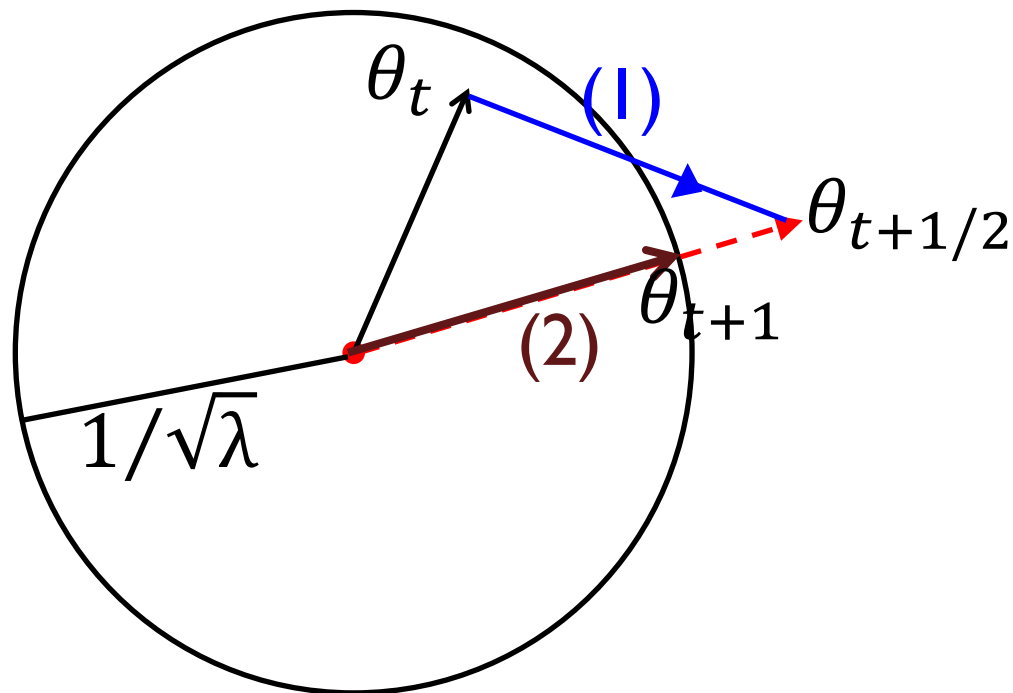
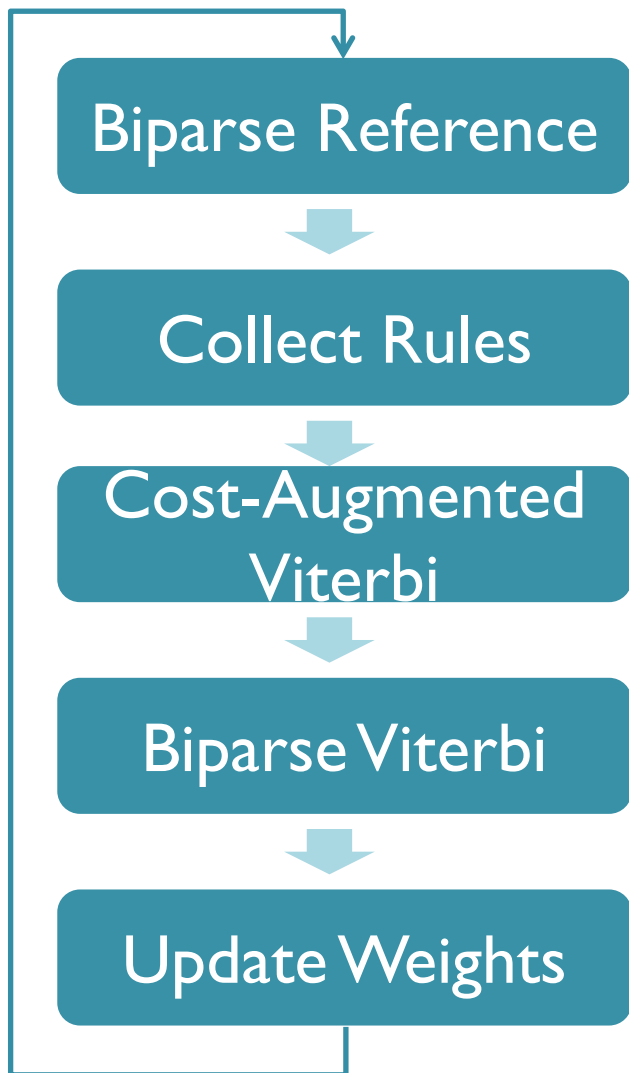
# Optimization

For each sent.



# Optimization

For each sent.



- (1) Sub-gradient: reference - viterbi
- (2) Projection: rescale weights

Shalev shwartz et al. (2007)

# Future Unsupervised Learning for SMT

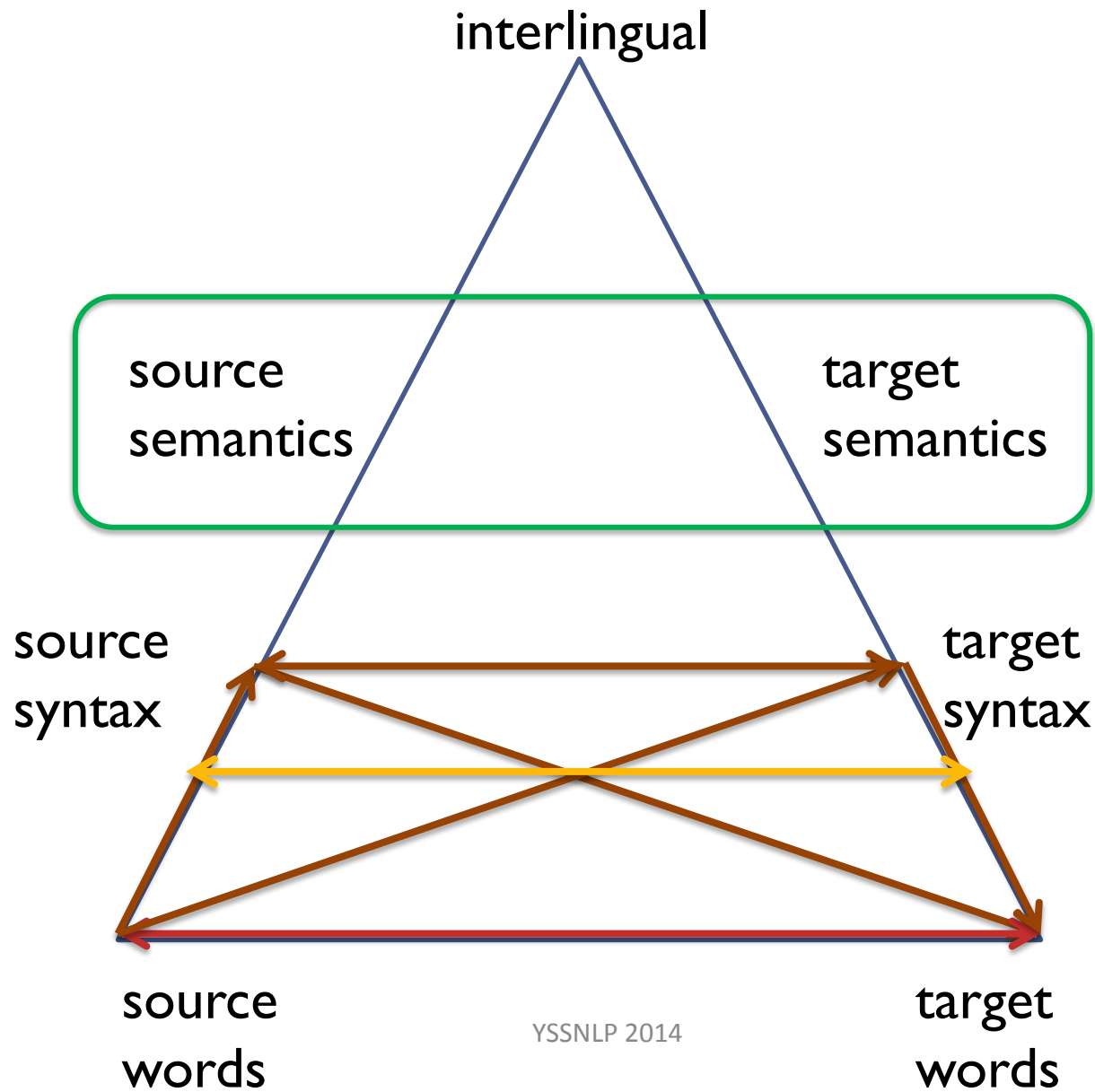
- ▶ Learning from big and unstructured data
- ▶ Learning tailored for new representations
- ▶ Never-ending learning



# SMT: Future Directions

- ▶ Representation
- ▶ Unsupervised learning
- ▶ **Semantics, discourse and SMT**

# Tower of Machine Translation



# Semantics and SMT: Multi-Level View

## ▶ Lexical semantics

- # word senses (Xiong and Zhang, 2014)

- # semantic roles (Xiong et al., 2012)

## ▶ Sentential semantics

- # compositional semantics

## ▶ Discourse-level semantics

- # cohesion (Xiong et al., 2013b, 2013c)

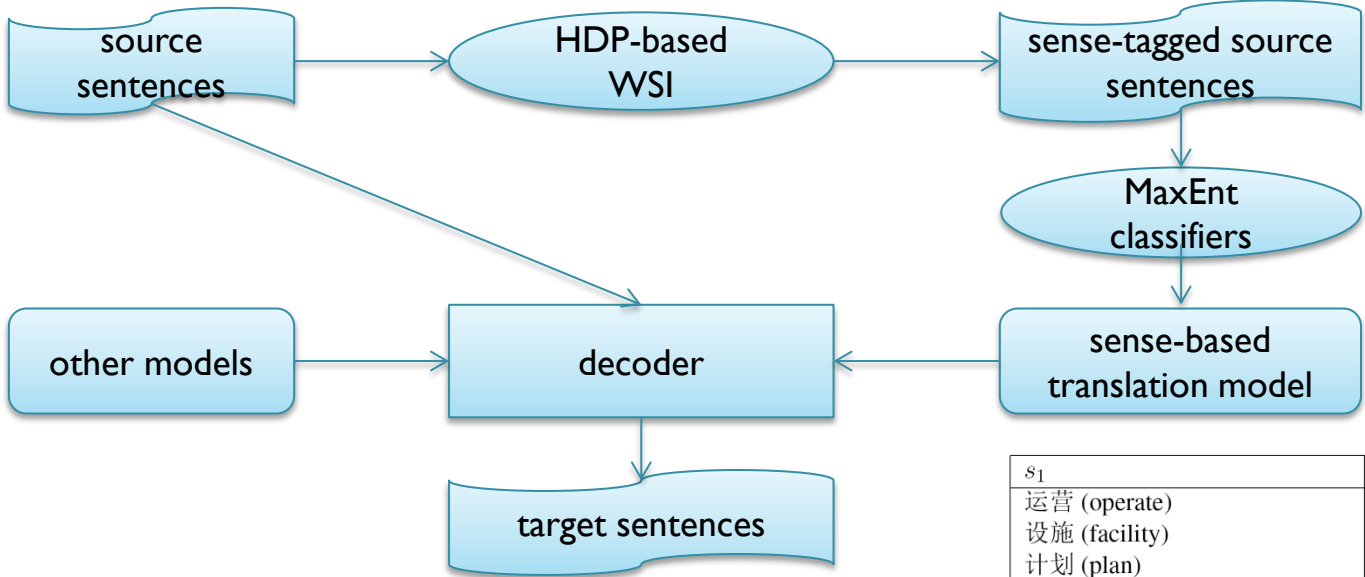
- # coherence (Xiong et al., 2013a)

# Semantics and SMT: Multi-Level View

- ▶ Lexical semantics
  - # word senses (Xiong and Zhang, 2014)
  - # semantic roles (Xiong et al., 2012)
- ▶ Sentential semantics
  - # compositional semantics
- ▶ Discourse-level semantics
  - # cohesion (Xiong et al., 2013b, 2013c)
  - # coherence (Xiong et al., 2013a)

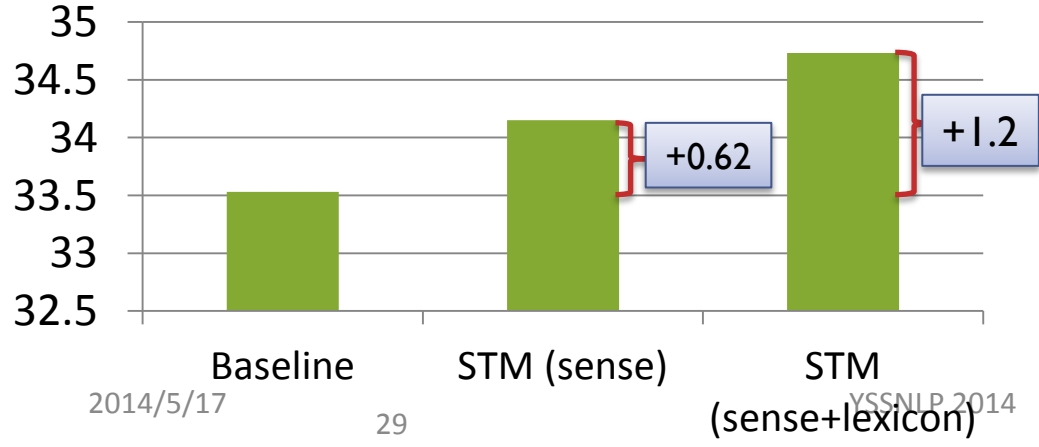
# Hidden Word Senses for SMT:

## Xiong and Zhang 2014



More info:  
A Sense-Based  
Translation  
Model for SMT

**BLEU-4**



s <sub>1</sub>	s <sub>2</sub>	s <sub>3</sub>
运营 (operate)	运营 (operate)	运营 (operate)
设施 (facility)	卫星 (satellite)	市场 (market)
计划 (plan)	系统 (system)	企业 (enterprise)
基础 (foundation)	国家 (country)	竞争 (competition)
项目 (project)	提供 (supply)	资产 (assets)
公司 (company)	国际 (inter-nation)	利润 (profit)
结构 (structure)	机构 (institution)	造成 (cause)
服务 (service)	进行 (proceed)	费用 (cost)
组织 (organization)	中心 (center)	资金 (capital)
提供 (supply)	合作 (cooperate)	业务 (business)
s <sub>4</sub>	s <sub>5</sub>	s <sub>6</sub>
费用 (cost)	城市 (city)	处于 (lie)
股价 (share price)	处理 (process)	拍照 (photograph)
27000	自来水 (tap-water)	119
科索沃 (Kosovo)	工厂 (factory)	DPRK
额外 (extra)	汽车 (car)	保险 (insurance)
工资 (wage)	铁路 (railway)	超支 (overspend)
美元 (dollar)	污水 (sewage)	地位 (position)
商业 (commerce)	办事处 (office)	经济 (economy)
收入 (income)	保本 (break-even)	竞争者 (competitor)
铁路局 (railway administration)	部件 (component)	平衡 (balance)

# Semantics and SMT: Multi-Level View

## ▶ Lexical semantics

- # word senses (Xiong and Zhang, 2014)

- # semantic roles (Xiong et al., 2012)

## ▶ Sentential semantics

- # compositional semantics

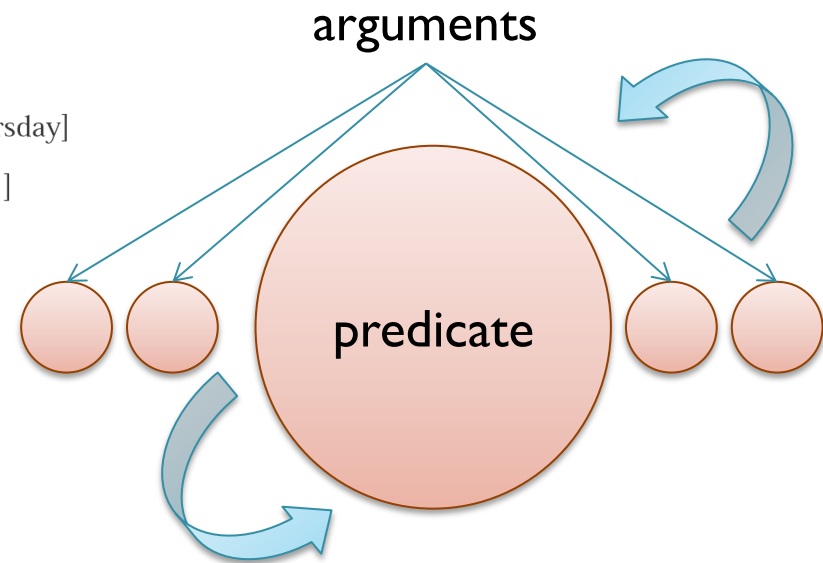
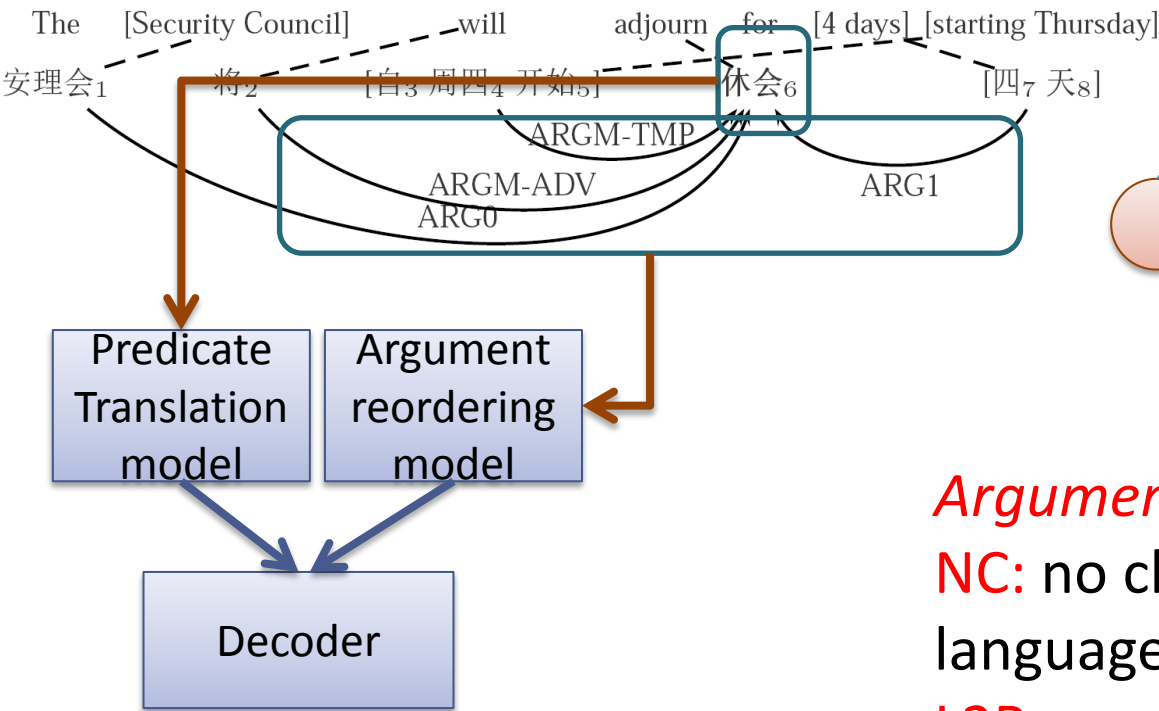
## ▶ Discourse-level semantics

- # cohesion (Xiong et al., 2013b, 2013c)

- # coherence (Xiong et al., 2013a)

# Semantic Roles for SMT:

## Xiong et al., 2012



**Argument reordering category:**

**NC:** no change across languages

**L2R:** moving from the left side of its predicate to the right side after translation

**R2L:** moving from the right side of predicate to the left side

Reordering Category	Percent
NC	82.43%
L2R	11.19%
R2L	6.38%

# Semantics and SMT: Multi-Level View

## ▶ Lexical semantics

- ⌘ word senses (Xiong and Zhang, 2014)

- ⌘ semantic roles (Xiong et al., 2012)

## ▶ Sentential semantics

- ⌘ compositional semantics

## ▶ Discourse-level semantics

- ⌘ cohesion (Xiong et al., 2013b, 2013c)

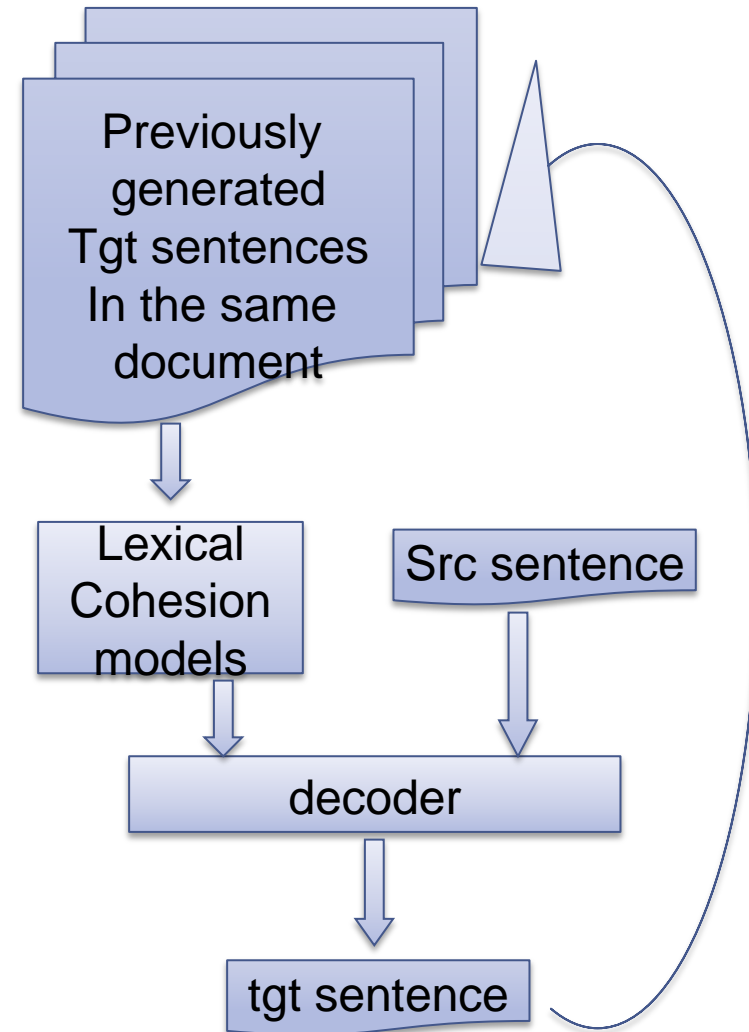
- ⌘ coherence (Xiong et al., 2013a)



# Discourse-Level Cohesion:

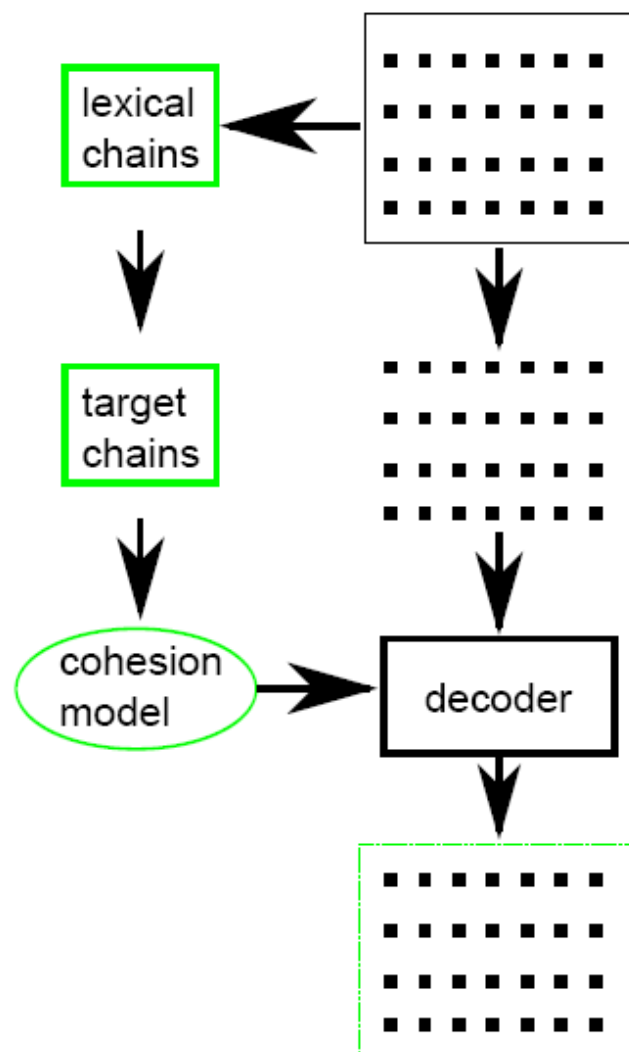
Xiong et al., 2013b

- ▶ **Direct reward model:** rewards translation hypotheses whenever lexical cohesion items occur in them
- ▶ **Conditional probability model:** measures the appropriateness of using lexical cohesion items
- ▶ **Mutual information trigger model:** estimates the strength of association between items in a lexical cohesion relation



# Discourse-Level Cohesion:

Xiong et al., 2013c



## Source chains generation

Compute lexical chains for each source document to be translated.

## Target chains projection

Project the computed source lexical chains onto the corresponding target document via maximum entropy classifiers.

## Cohesion model incorporation

Incorporate lexical cohesion into the target document translation via cohesion models built on the projected target lexical chains.

# Semantics and SMT: Multi-Level View

## ▶ Lexical semantics

- # word senses (Xiong and Zhang, 2014)
- # semantic roles (Xiong et al., 2012)

## ▶ Sentential semantics

- # compositional semantics

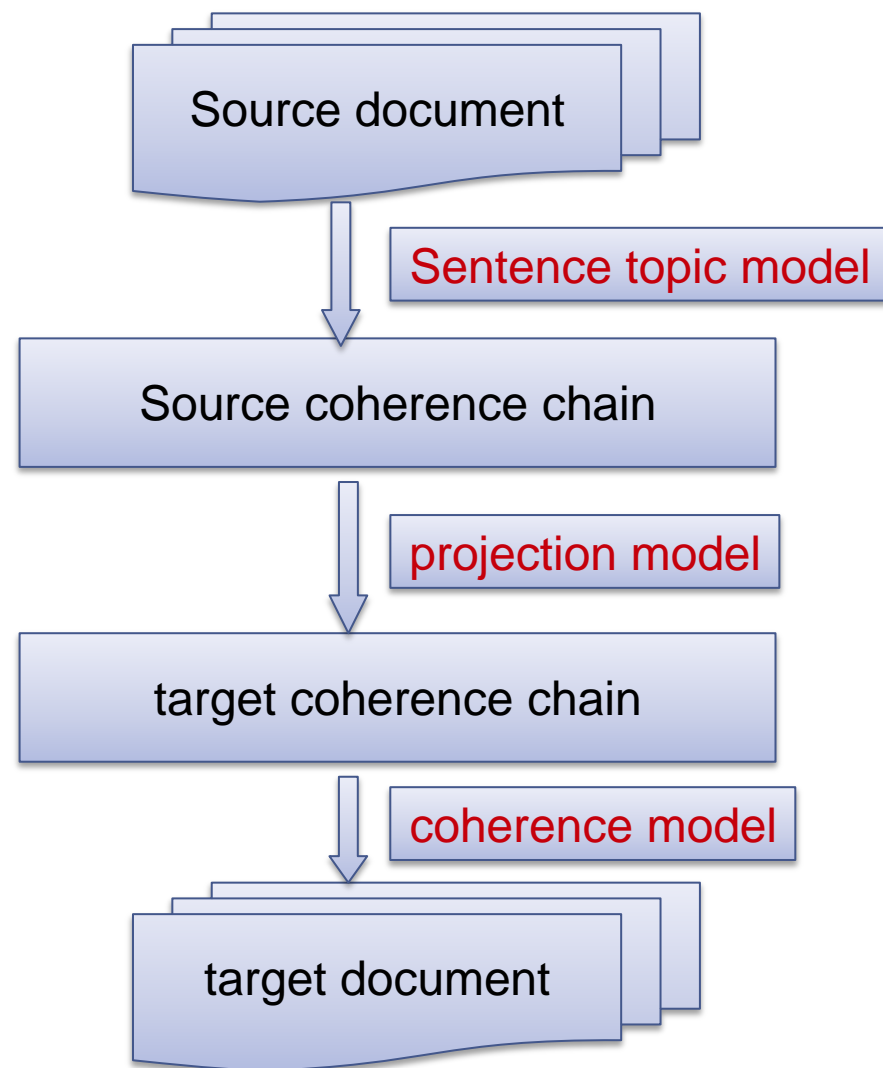
## ▶ Discourse-level semantics

- # cohesion (Xiong et al., 2013b, 2013c)
- # coherence (Xiong et al., 2013a)

# Discourse-Level Coherence:

Xiong and Zhang, 2013

- ▶ Model the sense continuity as a continuous sentence topic transition
- ▶ Characterize a document as a coherence chain of sentence topics
- ▶ Project source coherence chain to target documents



# Welcome to our ACL 2014 Tutorial

## Semantics, Discourse and SMT

# Tutorial Outline

- ▶ SMT Overall Review (30 minutes)
  - ▣ SMT architecture
  - ▣ phrase- and syntax-based SMT
- ▶ Semantics and SMT (1 hour and 15 minutes)
  - ▣ Brief introduction of semantics
  - ▣ Lexical semantics for SMT
  - ▣ Semantic representations in SMT
  - ▣ Semantically Motivated Evaluation
  - ▣ Advanced topics: deep semantic learning for SMT
  - ▣ Future directions
- ▶ Discourse and SMT (1 hour and 15 minutes)
  - ▣ Introduction of discourse: linguistics, computational and bilingual discourse
  - ▣ Discourse-based SMT: modeling, training, decoding and evaluation
  - ▣ Future directions

Special thanks to:

Xinyan Xiao  
Guosheng Ben  
Yang Ding

# References

- ▶ Xinyan Xiao, Deyi Xiong, Min Zhang, Qun Liu, Shouxun Lin. 2012. A Topic Similarity Model for Hierarchical Phrase-based Translation. ACL 2012.
- ▶ Xinyan Xiao, Deyi Xiong, Yang Liu, Qun Liu, Shouxun Lin. 2012. Unsupervised Discriminative Induction of Synchronous Grammar for Machine Translation. COLING 2012.
- ▶ Xinyan Xiao, Deyi Xiong. 2013 Max-Margin Synchronous Grammar Induction for Machine Translation. EMNLP 2013.
- ▶ Deyi Xiong, Min Zhang, Haizhou Li. 2012. Modeling the Translation of Predicate-Argument Structure for SMT. ACL 2012.



# References

- ▶ Deyi Xiong, Min Zhang. 2013a. A Topic-Based Coherence Model for Statistical Machine Translation. AACL 2013.
- ▶ Deyi Xiong, Guosheng Ben, Min Zhang, Yajuan Lv, Qun Liu. 2013b. Modeling Lexical Cohesion for Document-Level Machine Translation. IJCAI 2013.
- ▶ Deyi Xiong, Yang Ding, Min Zhang, Chew Lim Tan. 2013c. Lexical Chain Based Cohesion Models for Document-Level Statistical Machine Translation. EMNLP 2013.
- ▶ Deyi Xiong and Min Zhang. 2014. A Sense-Based Translation Model for Statistical Machine Translation. ACL 2014.

**Thank you!**