

ACL-08: HLT  
Columbus, Ohio

*A Linguistically Annotated  
Reordering Model for BTG-based  
Statistical Machine Translation*

*Deyi Xiong, Min Zhang, Aiti Aw, Haizhou Li  
Human Language Technology  
Institute for Infocomm Research  
dyxiong@i2r.a-star.edu.sg*

# Outline

- Previous work
- Baseline system
- The linguistically annotated reordering model
  - Model and its features
  - Annotation algorithm
  - Training and decoding
- experiments

# Previous work

- Flat reordering: Original BTG
  - Prior probability: can't capture reorderings between different phrases
- Boundary words based reordering model (Xiong et al. 2006)
  - Don't use any syntactic knowledge
- Reordering model using source-side syntactic features (Zhang et al. 2007)
  - Only reorder syntactic phrases

# Our work

- Still use boundary words as features
- Also use source-side syntactic knowledge for reordering
- Reorder any phrases: syntactic and non-syntactic
  - Linguistically annotated any phrases according to source-side parse tree
- Richer features: head words/tags beyond syntactic labels

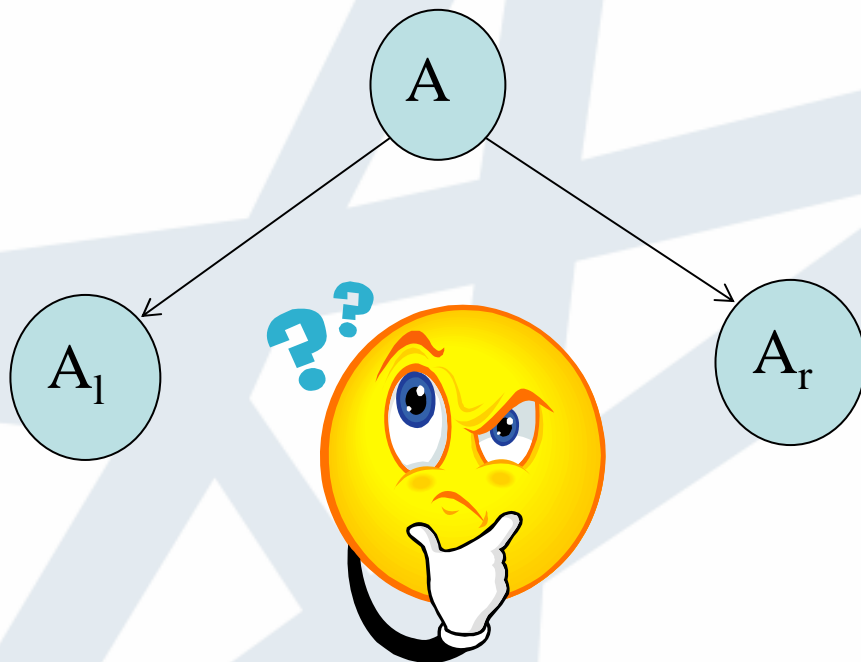
# Baseline system

- Using BTG rules
- log-linear phrase translation model
- Boundary words based reordering model
- CKY-style decoder

# Linguistically annotated reordering model (LAR)

- Model
- Annotation algorithm
- Training
- Decoding: how to use this model

# LAR model

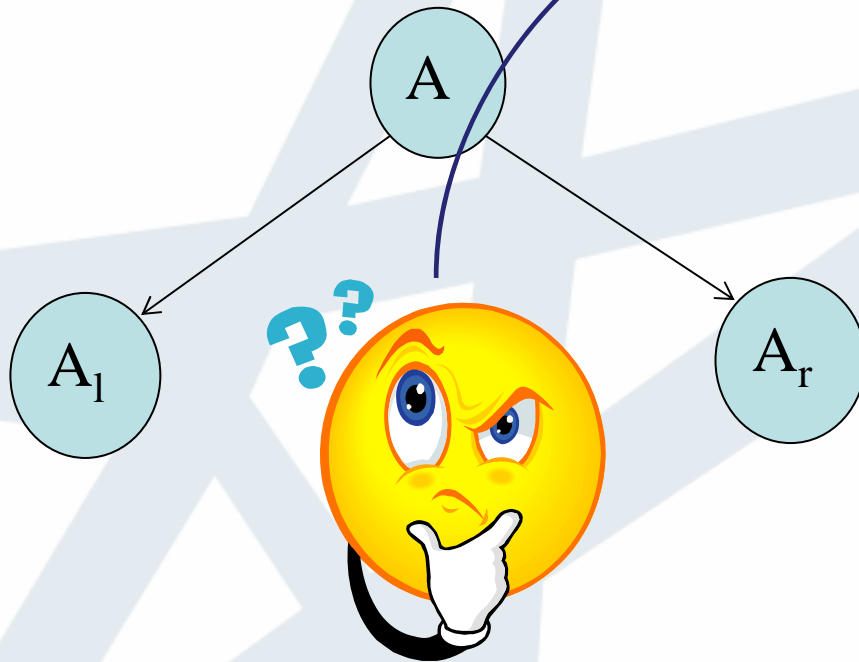


6/17/2008

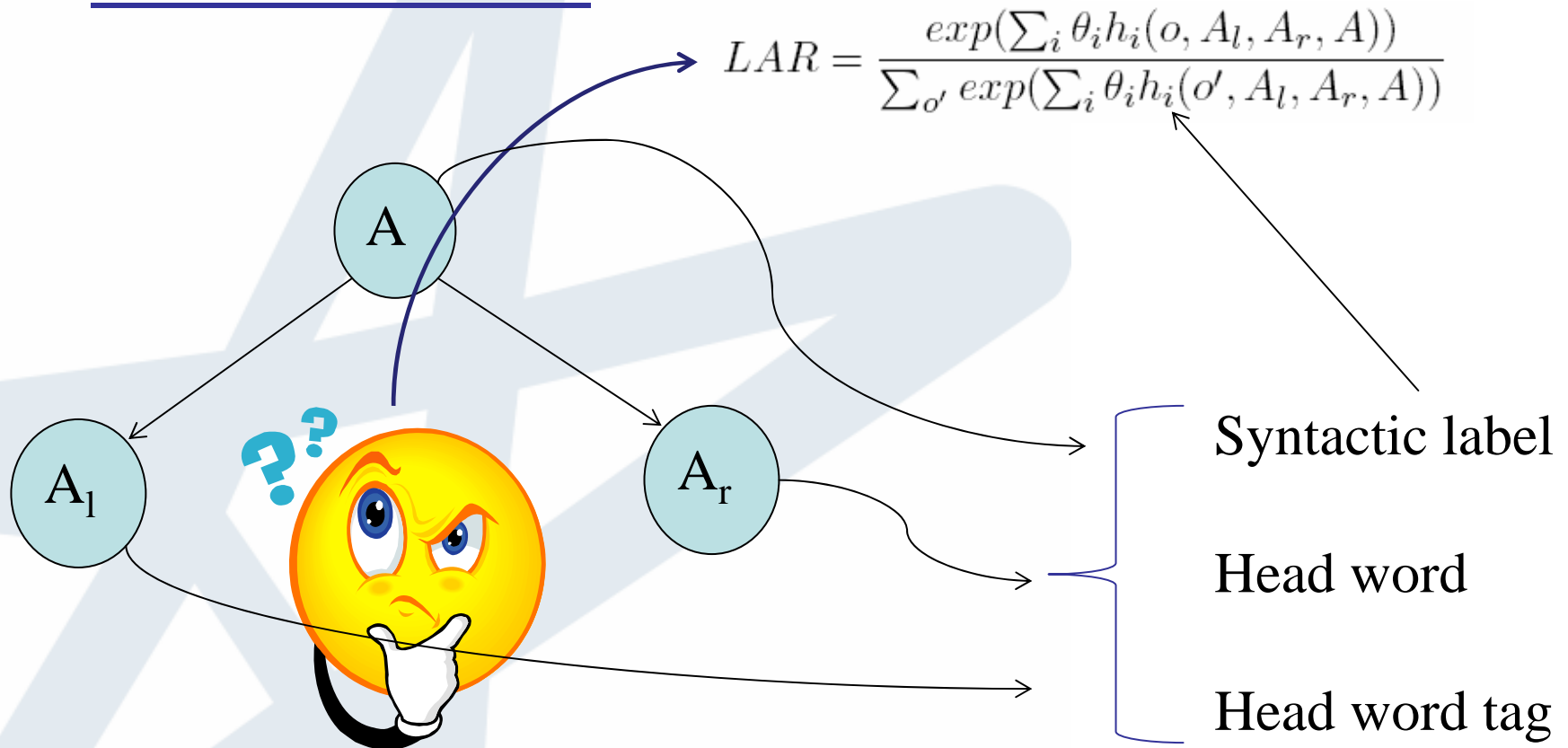
Deyi Xiong et al. ACL-08: HLT,  
Columbus, Ohio.

# LAR model

$$LAR = \frac{\exp(\sum_i \theta_i h_i(o, A_l, A_r, A))}{\sum_{o'} \exp(\sum_i \theta_i h_i(o', A_l, A_r, A))}$$



# LAR model

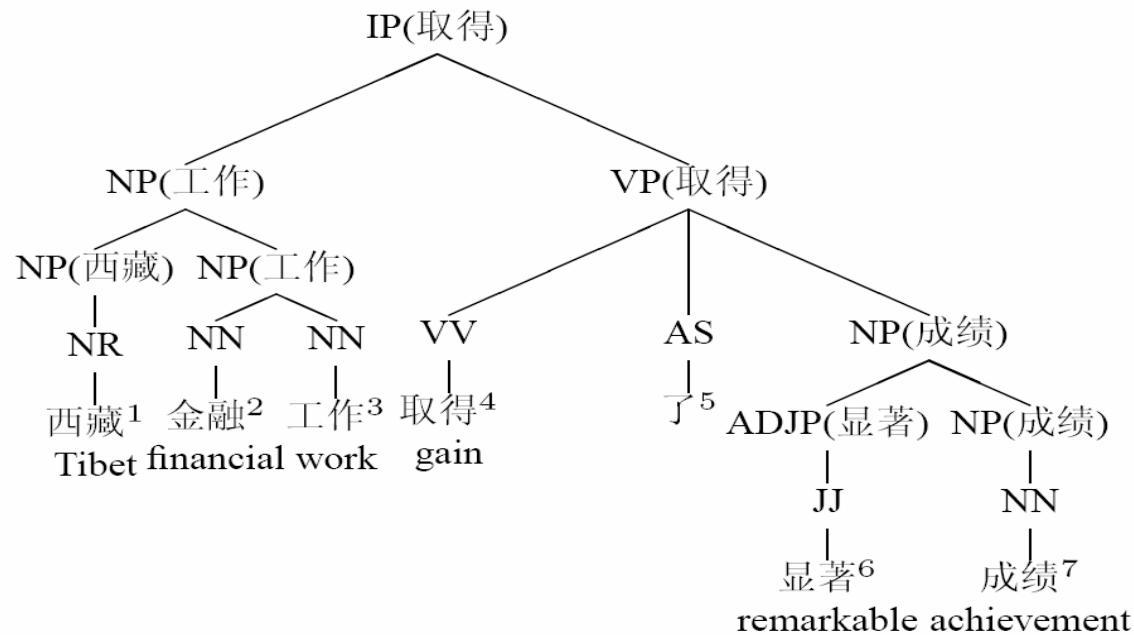


# Annotation algorithm

- Annotates any phrases (syntactic or non-syntactic) according to source-side parse tree

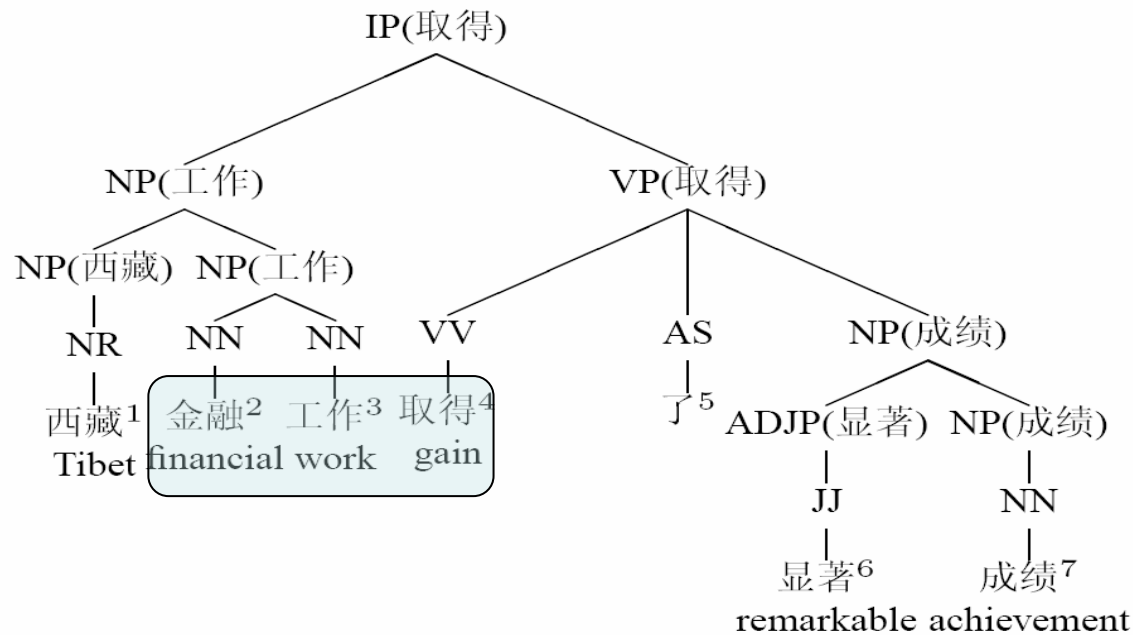
# Annotation algorithm

- Annotates any phrases (syntactic or non-syntactic) according to source-side parse tree



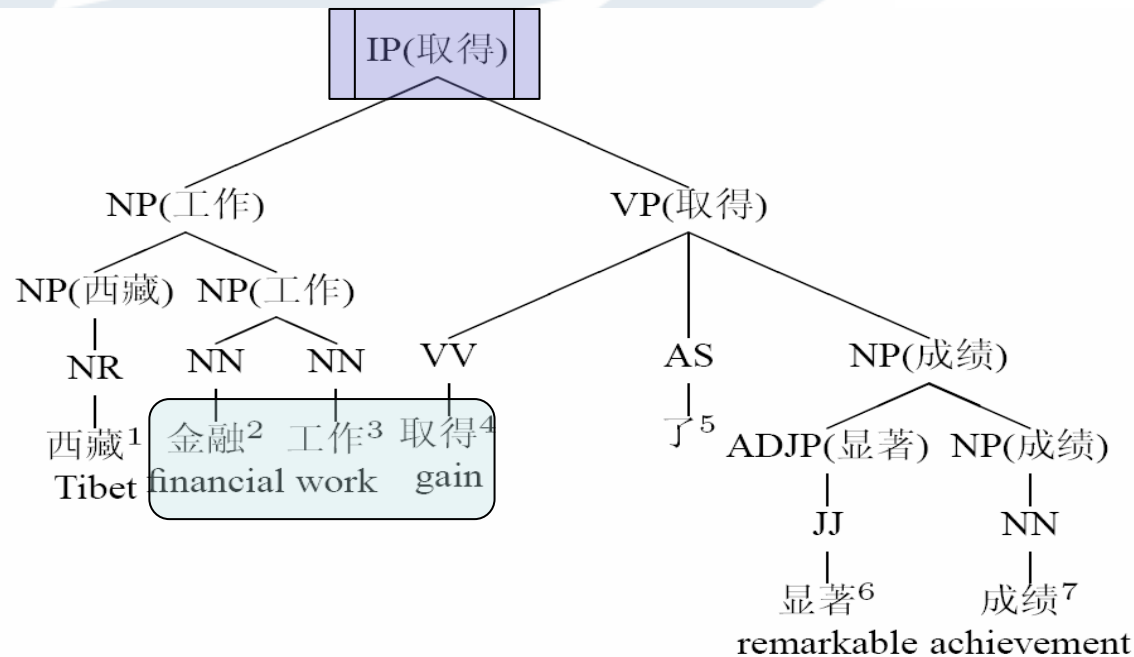
# Annotation algorithm

- Annotates any phrases (syntactic or non-syntactic) according to source-side parse tree



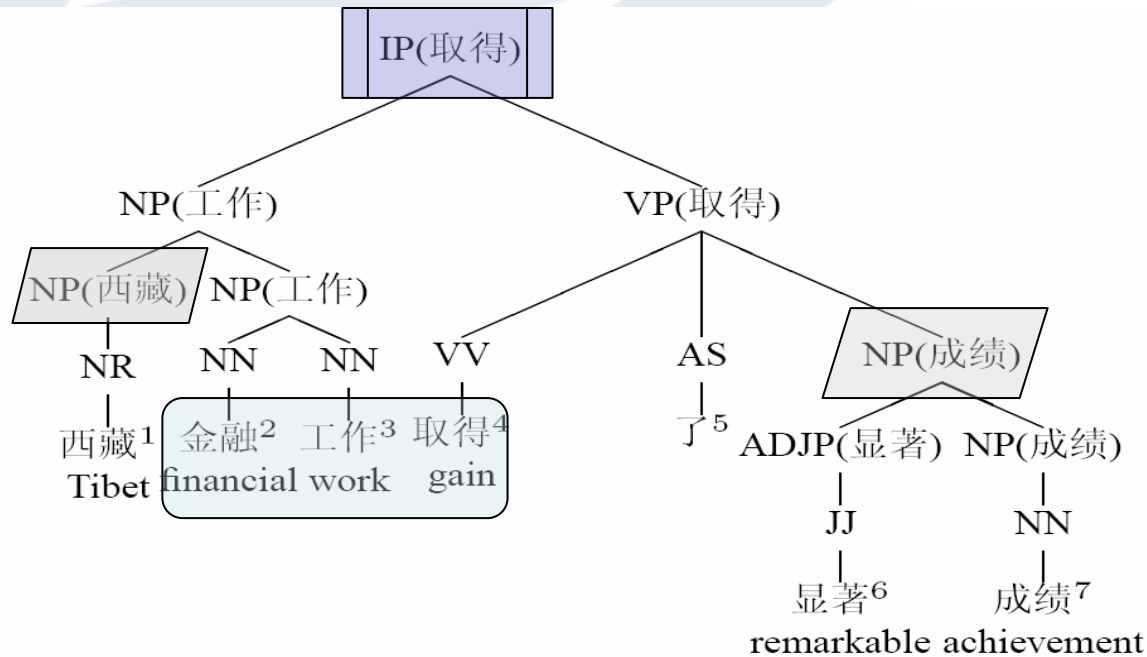
# Annotation algorithm

- Annotates any phrases (syntactic or non-syntactic) according to source-side parse tree



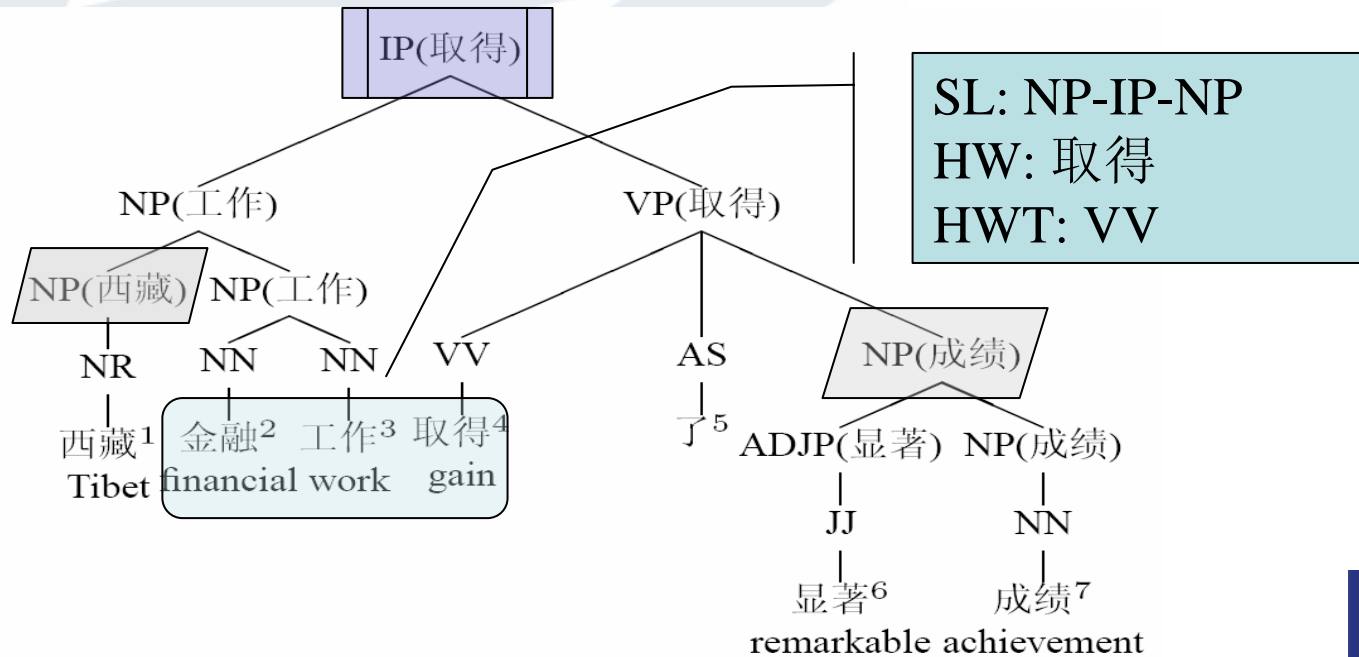
# Annotation algorithm

- Annotates any phrases (syntactic or non-syntactic) according to source-side parse tree



# Annotation algorithm

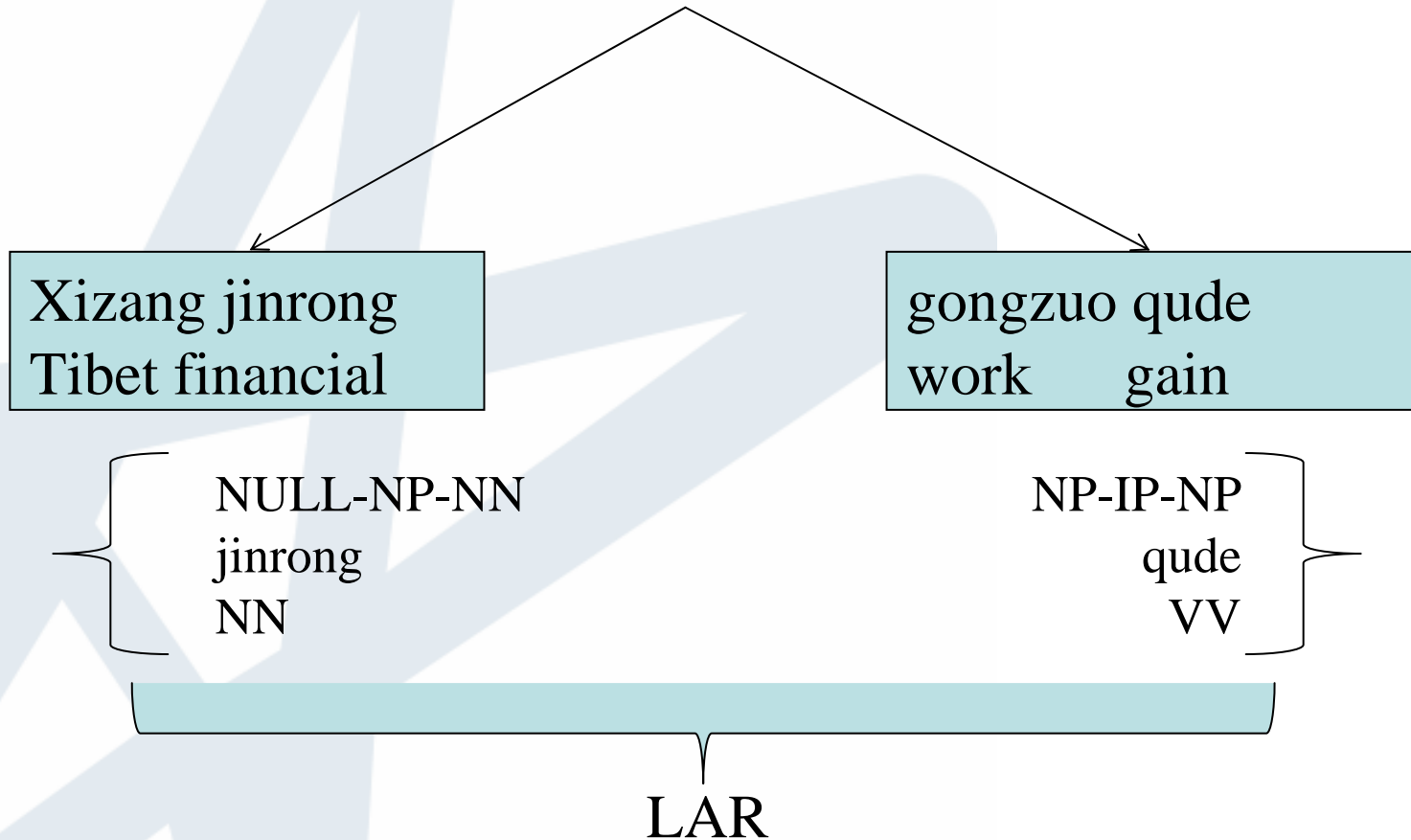
- Annotates any phrases (syntactic or non-syntactic) according to source-side parse tree



# LAR: training

- Extract reordering examples:  
annotated with syntactic label and  
head word/tag
- Generate features using linguistic  
annotations
- Estimate feature weights using MaxEnt  
tools

# Decoding



# Experiments

- Training data: FBIS
- 4-gram LM: trained on 181.1M words corpus
- Test set: NIST MT-05, Dev set: NIST MT-02
- We got:
  - 2.8M reordering examples
    - 114K features for BWR
    - 85K features for LAR

# Results

Reordering Configuration	BLEU (%)
BWR	24.97 $\pm$ 0.90
BWR + LAR (SL)	25.88 $\pm$ 0.95
BWR + LAR (+BNL)	26.27 $\pm$ 0.98
BWR + LAR (+BNL+HWT)	26.52 $\pm$ 0.96
Only allowed SPs reordering	25.12 $\pm$ 0.87

# Results

Reordering Configuration	BLEU (%)
BWR	24.97 ± 0.90
BWR + LAR (SL)	25.88 ± 0.95
BWR + LAR (+BNL)	26.27 ± 0.98
BWR + LAR (+BNL+HWT)	26.52 ± 0.96
Only allowed SPs reordering	25.12 ± 0.87

132 features

# Results

Reordering Configuration	BLEU (%)
BWR	24.97 ± 0.90
BWR + LAR (SL)	25.88 ± 0.95
BWR + LAR (+BNL)	26.27 ± 0.98
BWR + LAR (+BNL+HWT)	26.52 ± 0.96
Only allowed SPs reordering	25.12 ± 0.87

132 features

6.7K features

# Results

Reordering Configuration	BLEU (%)
BWR	24.97 ± 0.90
BWR + LAR (SL)	25.88 ± 0.95
BWR + LAR (+BNL)	26.27 ± 0.98
BWR + LAR (+BNL+HWT)	26.52 ± 0.96
Only allowed SPs reordering	25.12 ± 0.87

1.55 BLEU points

132 features

6.7K features

85K features

# Results

Reordering Configuration	BLEU (%)
BWR	24.97 ± 0.90
BWR + LAR (SL)	25.88 ± 0.95
BWR + LAR (+BNL)	26.27 ± 0.98
BWR + LAR (+BNL+HWT)	26.52 ± 0.96
Only allowed SPs reordering	25.12 ± 0.87

1.55 BLEU points

132 features

6.7K features

85K features



Thanks

6/17/2008

Deyi Xiong et al. ACL-08: HLT,  
Columbus, Ohio.

