Context-Aware Rule-Selection for Statistical Machine Translation

LIU Qun

CNGL @ Dublin City University &

ICT @ Chinese Academy of Sciences

Outline

- Introduction
- Context-Aware Rule-Selection
- CARS Application Examples
- Conclusion and Future Work

Linguistic Knowledge in SMT

- Used:
  - Morphology: segmentation
  - Syntax: Constituent, Dependency

So limited!

Linguistic Knowledge in SMT

- Unused：
  - Morphology: Inflection, Compound word
  - Syntax: Movement
  - Semantic: Preference, Semantic Role
  - Ontology
  - Discourse: Co-reference, Coherence, Topic Structure, Anaphora
  - Pragmatic: Sentiment, Intention, Situation…

So Much!

New SMT Paradigm?

- Word-based Translation
- Phrase-based Translation
- Syntax-based Translation
- ……
- Semantic-based Translation ???
- Discourse-based Translation ???
- ……

Problem

- Some of the translation problem may never be resolved without using certain kind of linguistic knowledge.

Example

- 10天前玛丽丢了一辆自行车。
- Mary lost **her** **bicycle** 10 days ago.

- 刚才警察来通知车找到了。
- Just now the police come to tell **her** that **her** **bicycle** was found.

- Need ontology：“自行车” is-kind-of “车”
- Need coreference resolution for insert “her”

Problem

- Some linguistic theory only have effect on very specific language phenomenon

- Building a new SMT paradigm on a certain linguistic knowledge (x-based translation)
  - high cost
  - usually lead to decrease of BLEU scores

Our Solution: CARS

Context-Aware Rule Selection

- Compatible to current log-linear SMT framework
- Easy to integration various linguistic knowledge to current SMT system
- Working locally rather than globally
- Effect!

Example：mouse

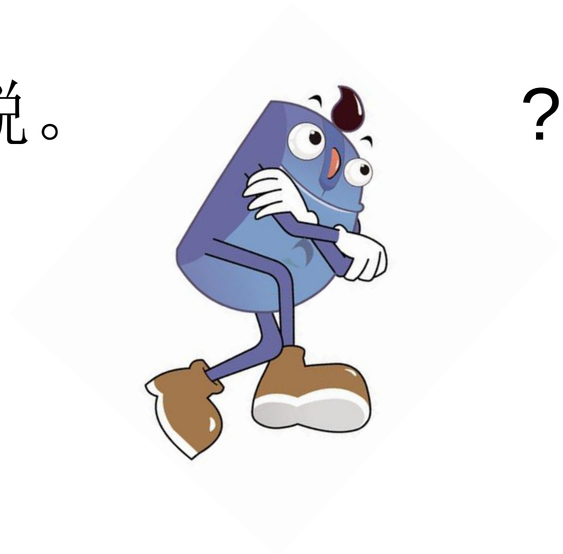
mouse	老鼠
	鼠标

- The **mouse** was found to have **escaped** two days later.
- 两天后发现这只**老鼠逃跑**了。

- The **mouse** was found **damaged** two days later.
- 两天后发现这只**鼠标坏**了。

Google Translate

- The **mouse** was found to have escaped two days later.
- **鼠标**两天后逃脱。 ???



- The **mouse** was found damaged two days later.
- 两天后发现损坏的**鼠标**。

Example：X_1 的 X_2

X_1 的 X_2	X_1 's X_2
	X_1 X_2
	X_2 of X_1
	X_2 of X_1 's

狐狸**的**尾巴

the fox**'s** tail

地球**的**环境

the environment **of** the earth

木头**的**桌子

wood table

小王**的**一个朋友

a friend **of** Xiao Wang**'s**

Google Translate

- 狐狸的尾巴
- Fox tail

- 地球的环境
- Earth’s environment

- 小王的一个朋友
- Wang a friend

- 木头的桌子
- Wood table

Notions

- **Language Expression:**

  an expression used in statistical translation model as a description of a piece of language surface form or certain language structure.

Notions

- **Translation Rule:**

  a mapping from a source language expression to a target language expression

- **Translation Rule Selection:**

  to select the best target language expression for a given source language expression, by giving a score to each candidate translation rule

Language Expression

- Word
- Phrase
- CFG Rule
- CFG Tree
- Dependency Rule
- Dependency Treelet
- String of Terminals and Non-T.
- ……

Translation Rules

Translation Models	Translation Rules
IBM Model 1-5	Word → Word (word translation table)
Phrase-based Model	Phrase → Phrase (phrase table)
Hierarchical Phrase-based Model	CFG Rule → CFG Rule
String-to-Dependency (Shen 08)	CFG Rule → CFG rule with Dep.
Tree-to-String Model	CFG Tree → String
String-to-Tree Model	String → CFG Tree
Dependency Model (Quirk 05)	Dep. Treelet → Dep. Treelet
Dependency Model (Xiong 06)	Dep. Treelet → String
Dependency Model (Xie 11)	Dep. Rule → String

Rule Selection

mouse	老鼠
	鼠标

X_1 的 X_2	X_1 X_2
	X_1 ’s X_2
	X_2 of X_1
	X_2 of X_1 ’s

Rule Selection

Given S, select rule from:

$$\begin{cases} r_1\colon S \rightarrow T_1 \\ r_2\colon S \rightarrow T_2 \\ \qquad\vdots \\ r_n\colon S \rightarrow T_n \end{cases}$$

Rule Selection by Probability

$$\hat{r} = \underset{r_i}{\operatorname{argmax}} P(r_i|S)$$

$$where: \sum_i P(r_i|S) = \sum_i P(T_i|S) = 1$$

Rule Selection by Probability

mouse	**老鼠**	**0.4**
	鼠标	**0.6**

X_1 的 X_2	**X_1 X_2**	**0.3**
	X_1 's X_2	**0.4**
	X_2 of X_1	**0.2**
	X_2 of X_1 's	**0.1**

Problem

- All probabilities for rule selection are static values trained from the training corpus.
- No context information is able to be used for rule selection.
- Language model and reordering model only help a little for rule selection.

Outline

- Introduction
- Context-Aware Rule-Selection
- Implementation of CARS
- Conclusion and Future Work

Motivation

Rule Selection

by

Dynamic Context Information

Context-Aware Rule Selection —— CARS Model

$$\mathrm{Score}(r_i|C,S)$$

$r_i: S \rightarrow T_i$: Translation Rule

C: Context

S: Source Expression

Note: CARS model is used as a feature of the log-linear model in SMT.

Probabilistic CARS Model

$$\text{Score}(S, C) = P(r_i | C, S)$$

$$where: \sum_i P(r_i | C, S) = 1$$

Note: As a feature of log-linear model, CARS model is not necessary to be a probability.

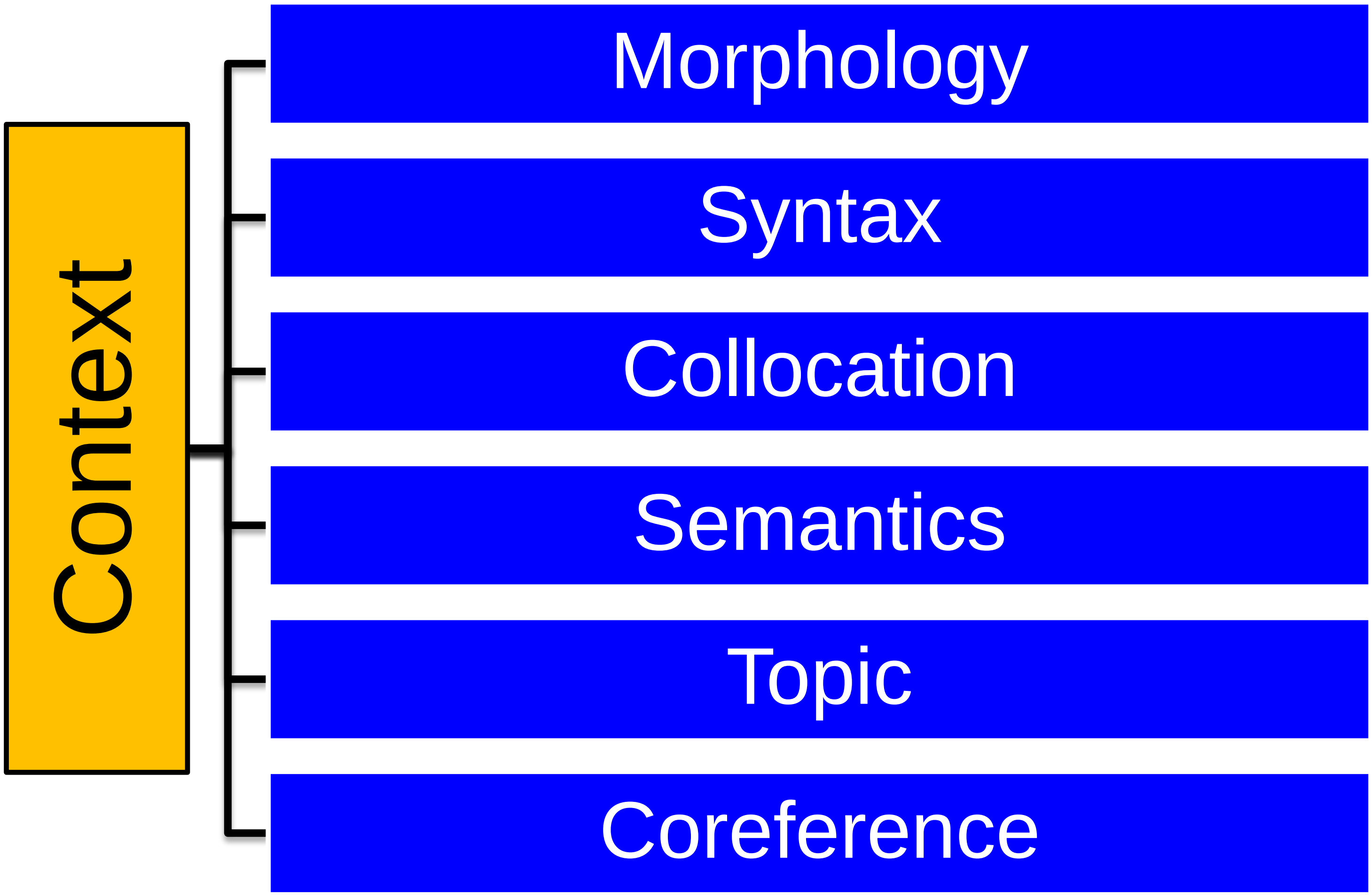
Discriminative CARS Model

$$P(r_i|S,C) = \frac{\exp(\sum_k \lambda_k h_k(r_i, C|S))}{\sum_{r_j} \exp(\sum_k \lambda_k h_k(r_j, C|S))}$$

$h_k(r, C)$: Context Features

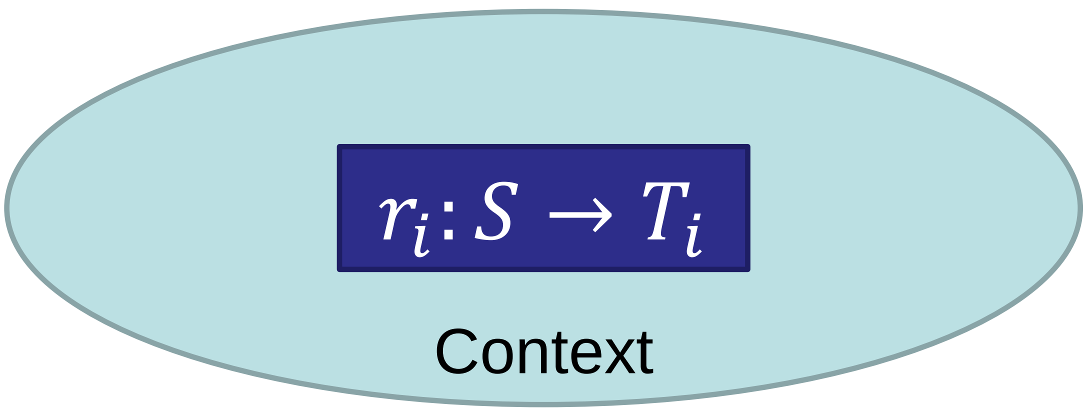
λ_k: Weights of Context Features

Context Features



Training CARS Model

- To training a CARS model, we need:
  - Count the number of the rules (as usual)
  - Reserve the context for each occurrence of the rule (new requirement)



Applicability of CARS Model

- CARS model may applicable only to part of the rules, for example:
  - only for lexicalized rules
  - only for un-lexicalized rules
  - only for verbs (SRL)
  - only for pronouns (Coreference)
  - only for to a single word (DE)
  - ……

CARS Utilization as a Feature

- An additional feature of CARS Utilization may be also necessary in log-linear model
  - To record the times of using CARS model in decoding
  - To balance between the rules using or not using CARS model
  - Not necessary if the CARS model is applicable to all rules

Outline

- Introduction
- Context-Aware Rule-Selection
- CARS Application Examples
- Conclusion and Future Work

CARS A

Deyi Xiong et al. COLING-ACL2006

Zhongjun He et al. COLING2008

CARS for Br

Qun Liu et al. EMNLP2008

CARS for H

Xinyan Xiao et al. ACL2012

CARS for

unpublished

CARS using To odel

CARS for Agglutinative Language Translation

CARS Application Examples

- CARS for Bracketing Transduction Grammar
- CARS for Hierarchical Phrase-based Model
- CARS for Tree-to-String Model
- CARS using Topic Model
- CARS for Agglutinative Language Translation

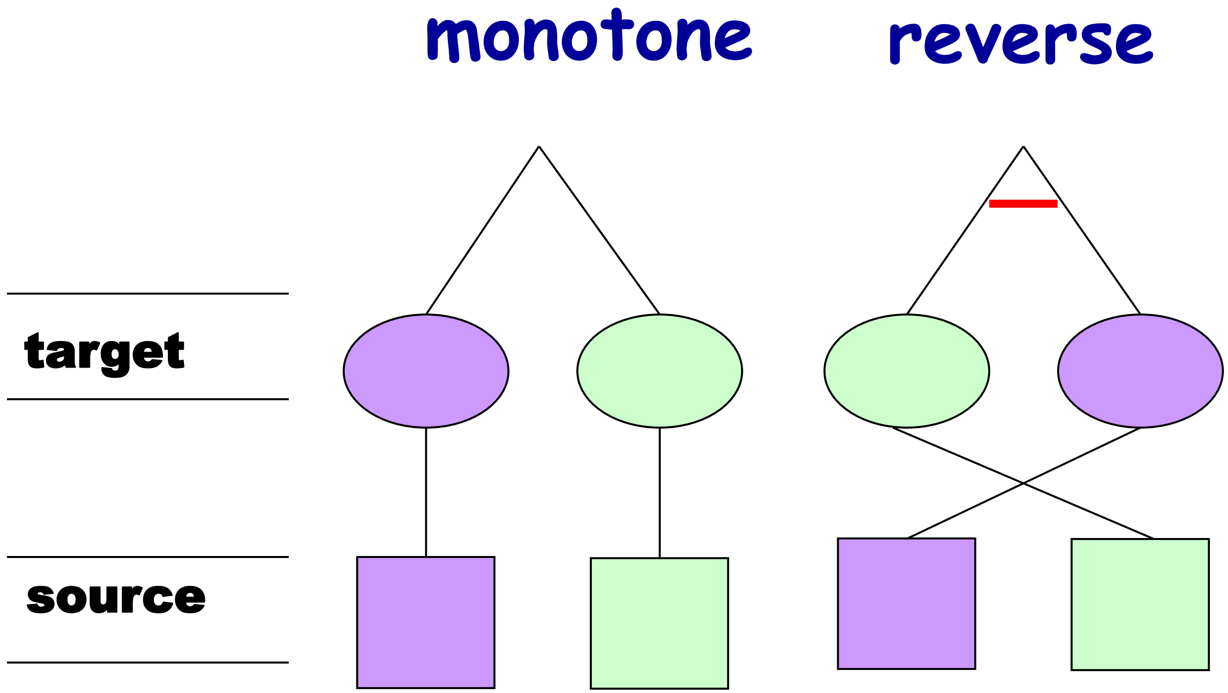
ITG: Inversion Transduction Grammar

(Wu, Dekai 1995)

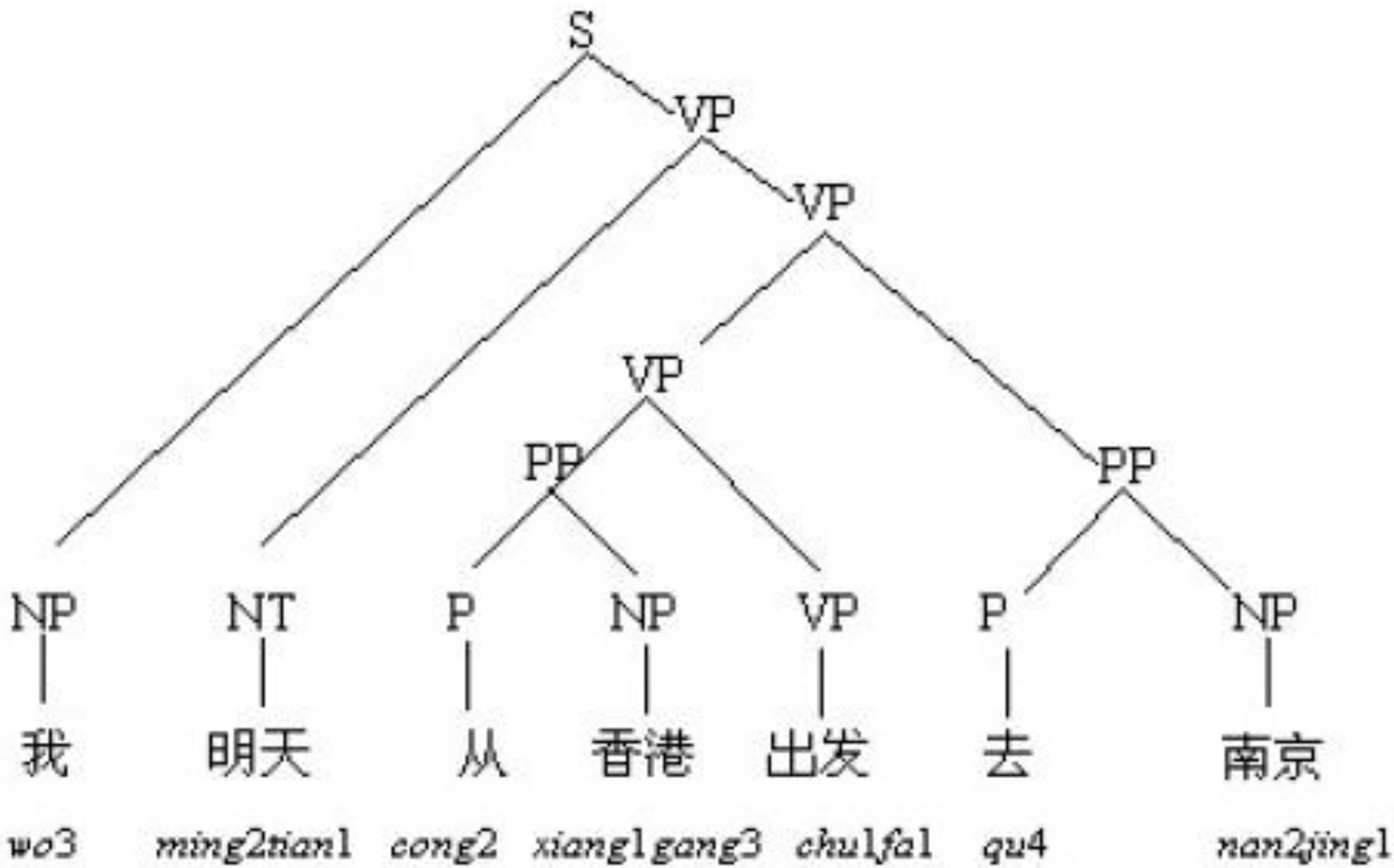
- Synchronized Grammar
- Binary Rules (CNF style)

ITG rules	Source	Target
$A \rightarrow [\,B\ C\,]$	$A \rightarrow BC$	$A \rightarrow BC$
$A \rightarrow \langle B\ C \rangle$	$A \rightarrow BC$	$A \rightarrow CB$
$A \rightarrow x/y$	$A \rightarrow x$	$A \rightarrow y$

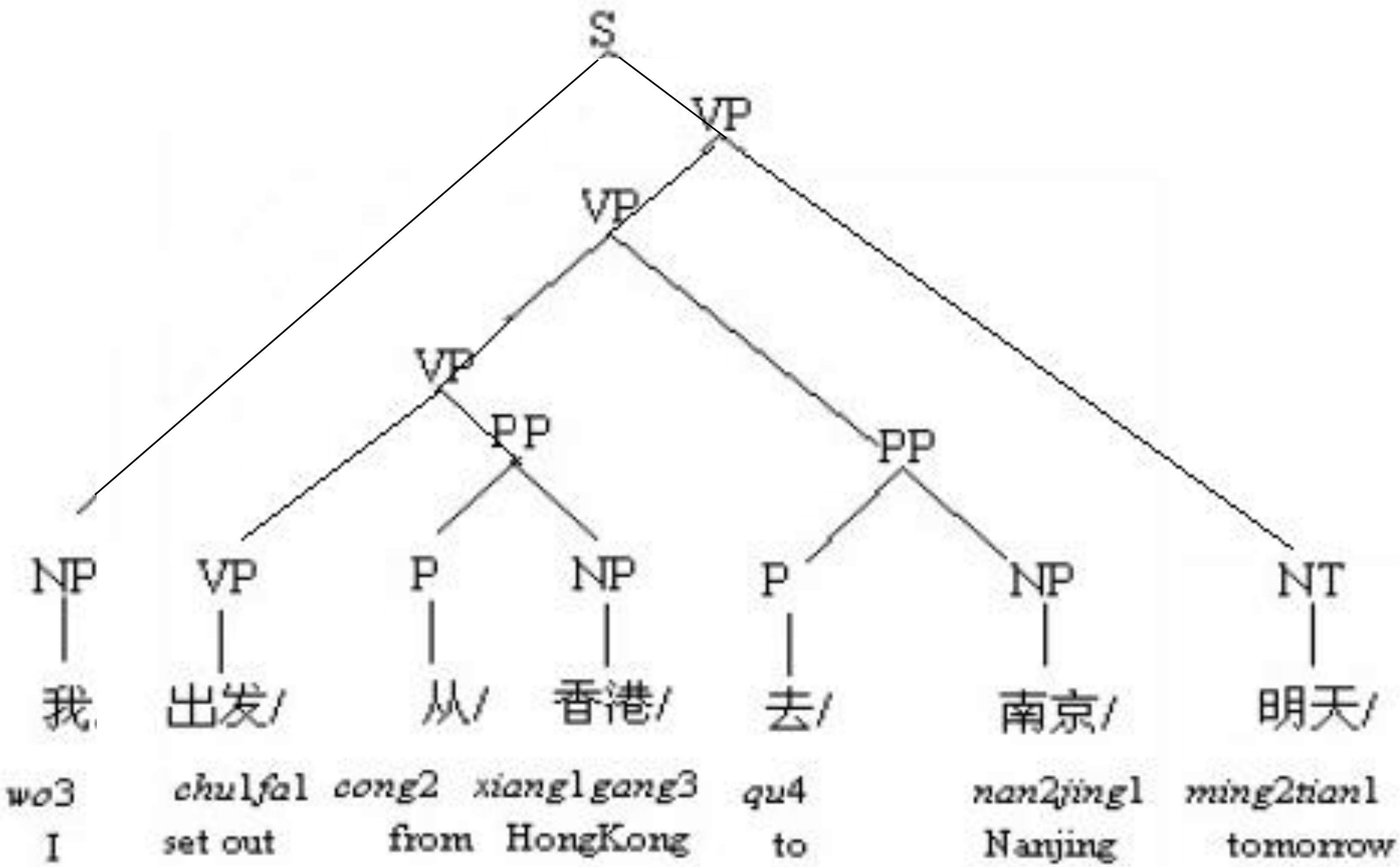
ITG Non-Terminal Rules



ITG Based Translation (1)



ITG Based Translation (2)



ITG Based Translation (3)

- Pros：
  - Recursive
  - Linguistic-style grammar
  - Limited search space
- Cons:
  - Need human annotated bi-lingual corpus for training

BTG: Bracketing Transduction Grammar

- BTG:

  A simplified ITG with only one non-terminal

- Only two non-terminal rules:

  $X \rightarrow [\, X_1 \;\; X_2 \,]$ (monotone rule)

  $X \rightarrow \langle\, X_1 \;\; X_2 \,\rangle$ (reverse rule)

Stochastic BTG

(Wu, Dekai 96)

- Static Rule Selection
- Only one parameter for non-terminal rules

  $X \rightarrow [X_1 \ X_2] $: p(monotone rule)=0.7

  $X \rightarrow \langle X_1 \ X_2 \rangle$: p(reverse rule)=0.3
- Too week discriminability
- Our Approach: CARS

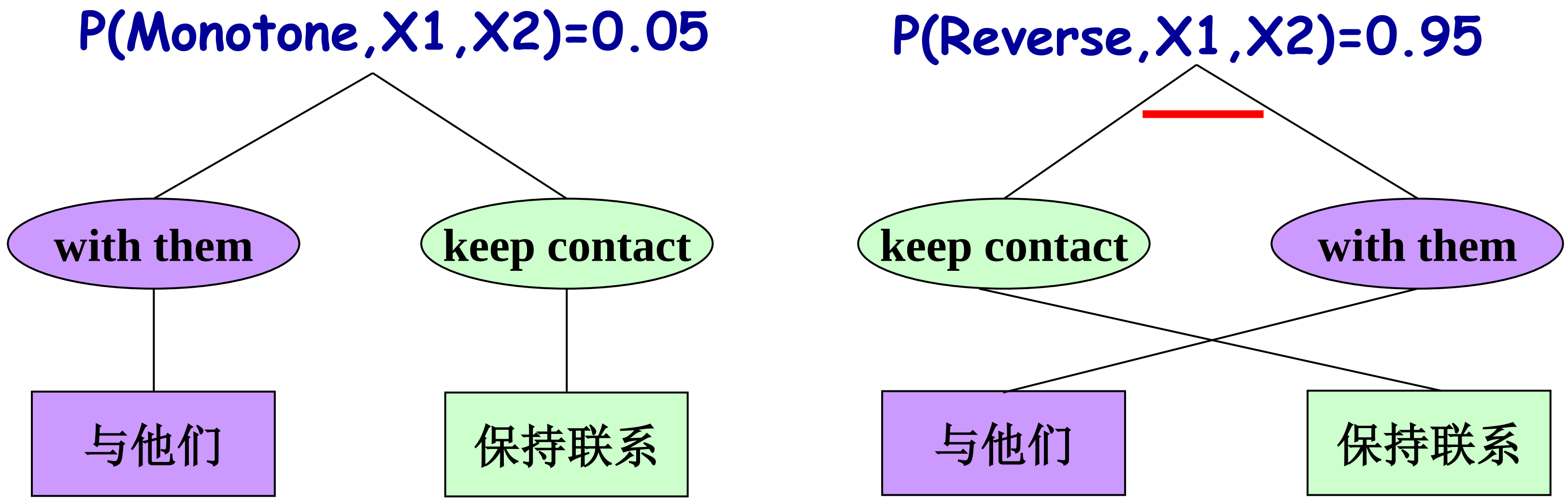
Our Method

- Given bilingual phrase X_1 and X_2

  X_1= “with them◇与他们”

  X_2= “keep contact◇保持联系”

- Calculate the probabilities using X_1 and X_2:



Maximum-Entropy BTG

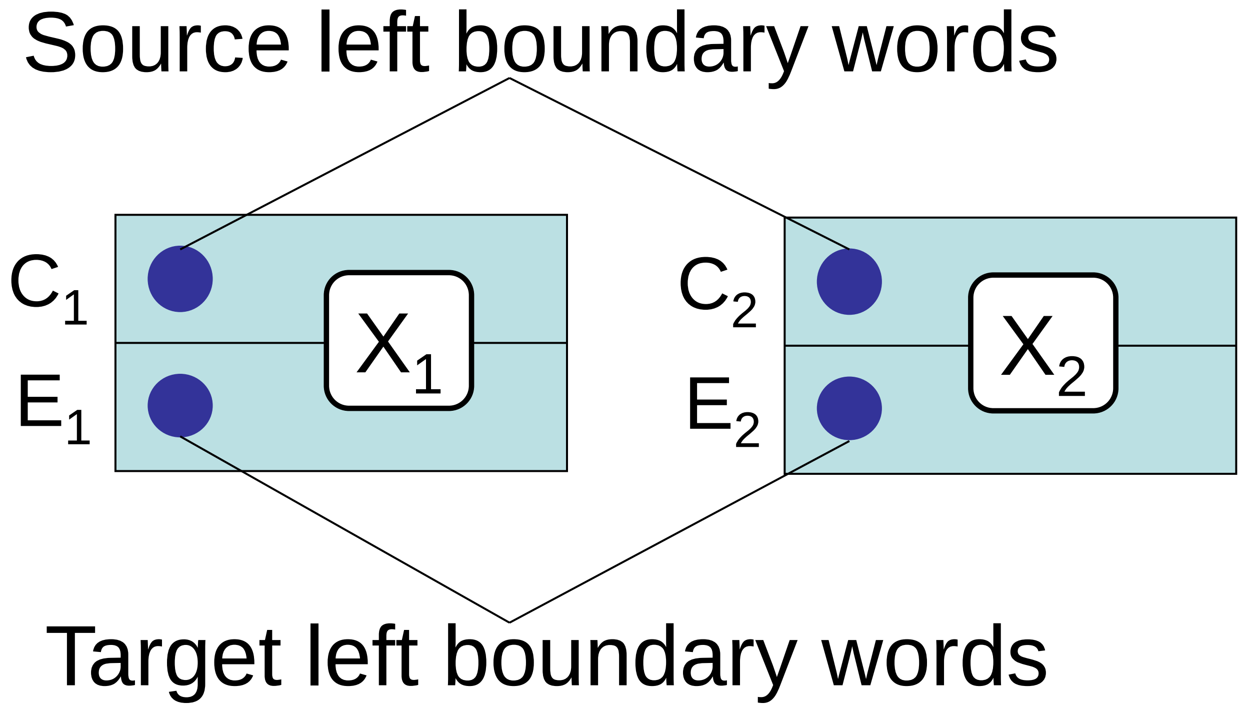
- Modeling: Maximum Entropy

$$\Omega = p_\theta(o \mid X_1, X_2) = \frac{exp(\sum_i \theta_i h_i(o, X_1, X_2))}{\sum_{o'} exp(\sum_i \theta_i h_i(o', X_1, X_2))}$$

$$h_i(o, X_1, X_2) = \begin{cases} 1 & \text{if } f(X_1, X_2) = \text{True}, \quad o = 0 \\ 0 & \text{otherwise} \end{cases}$$

$$0 \in \{\text{monotone}, \text{reverse}\}$$

Features



We ONLY use monolingual or bilingual left boundary words as features

Feature Templates

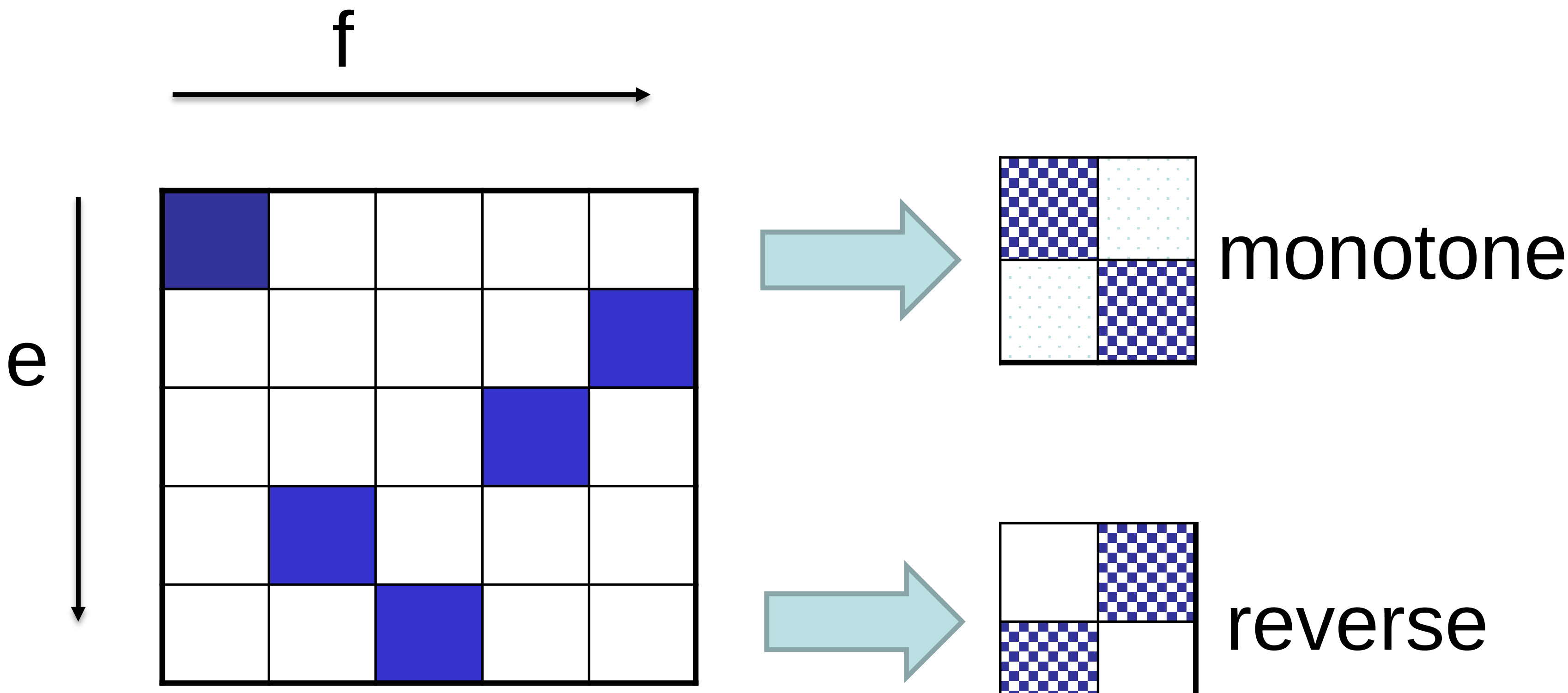
X_1= “with them◇与他们”

X_2= “keep contact◇保持联系”

C1	C1=与
C2	C2=保持
E1	E1=with
E2	E2=keep
C1C2	C1=与 & C2=保持
C1E1	C1=与 & E1=with
C2E2	C2=保持 & E2=keep
E1E2	E1=with & E2=keep

Training Samples Extraction

Word Alignment



Experiment Result

Systems	NIST MT-05	IWSLT-04
monotone	20.1 ± 0.8	37.8 ± 3.2
NONE	19.6 ± 0.8	36.3 ± 2.9
Distortion	20.9 ± 0.8	38.8 ± 3.0
Flat	20.5 ± 0.8	38.7 ± 2.8
Pharaoh	20.8 ± 0.8	38.9 ± 3.3
MaxEnt (lex)	*22.0* ± 0.8	*42.4* ± 3.3
MaxEnt (lex + col)	**22.2** ± 0.8	**42.8** ± 3.3

Summary

- We proposed MEBTG to compute the probability of two BTG non-terminal rules.
- Only boundary word features are used in MEBTG model.
- MEBTG model is very effective as a reordering model for phrase-based translation.
- A lot of citations and follow-up works.

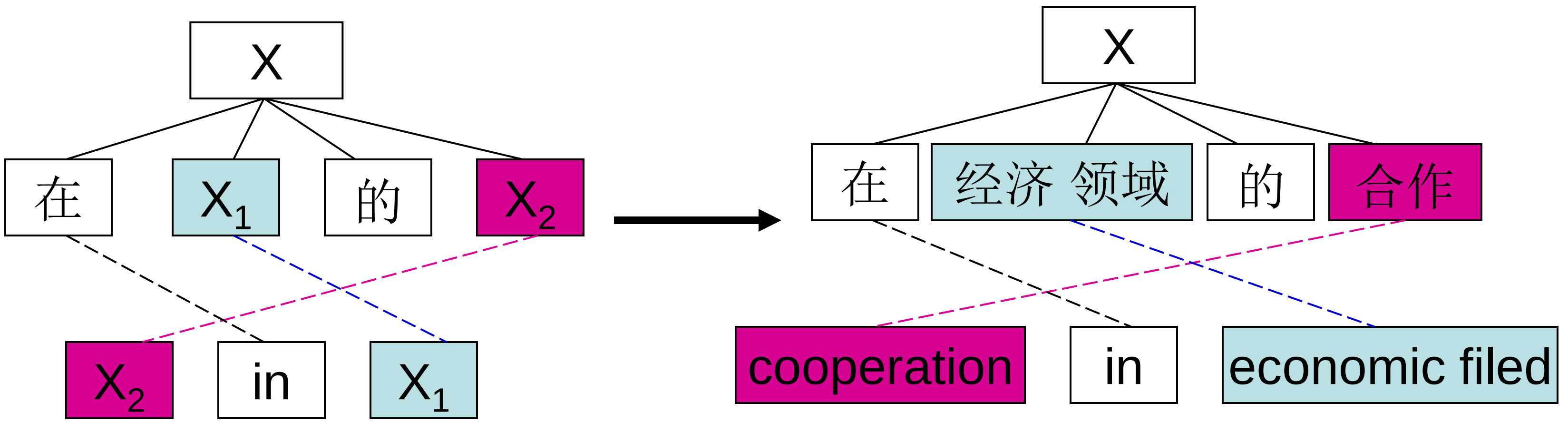
CARS Application Examples

- CARS for Bracketing Transduction Grammar
- CARS for Hierarchical Phrase-based Model
- CARS for Tree-to-String Model
- CARS using Topic Model
- CARS for Agglutinative Language Translation

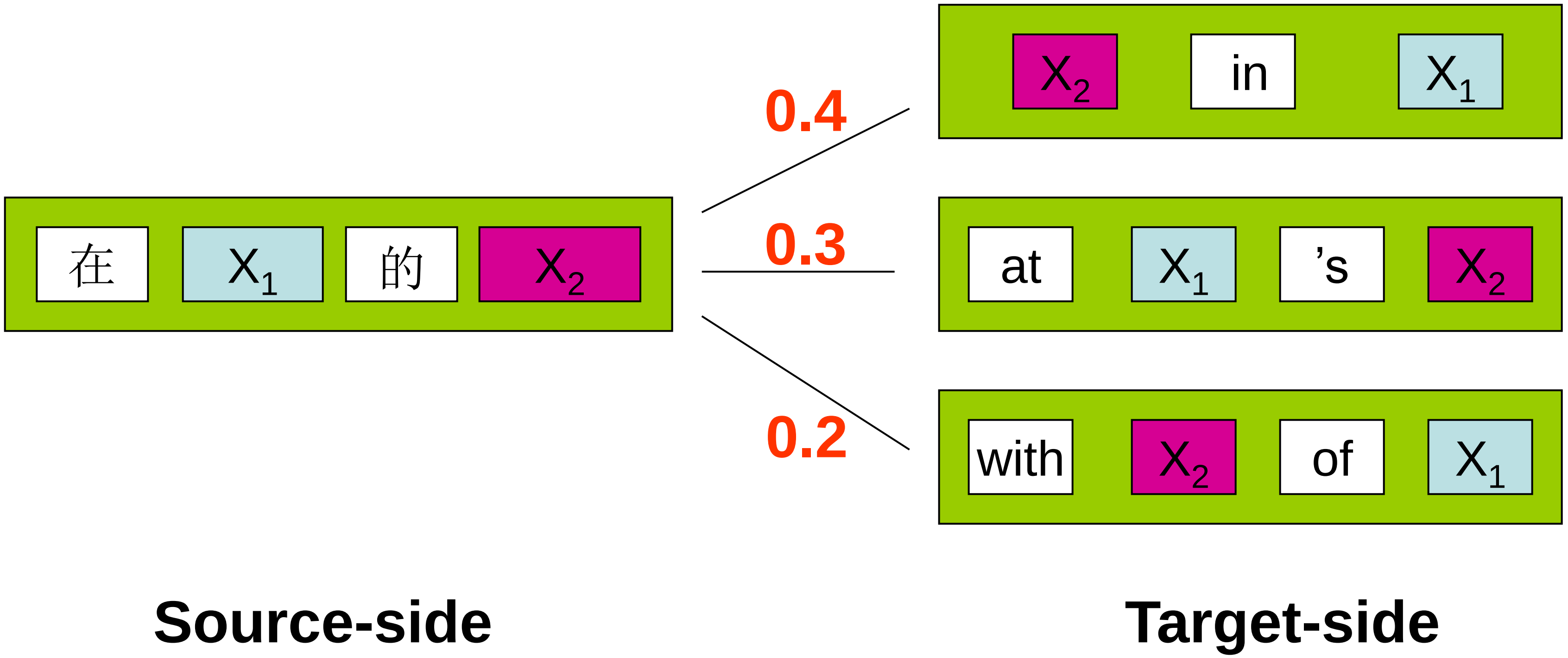
Hierarchical Phrase-Based Model

David Chiang. ACL2005

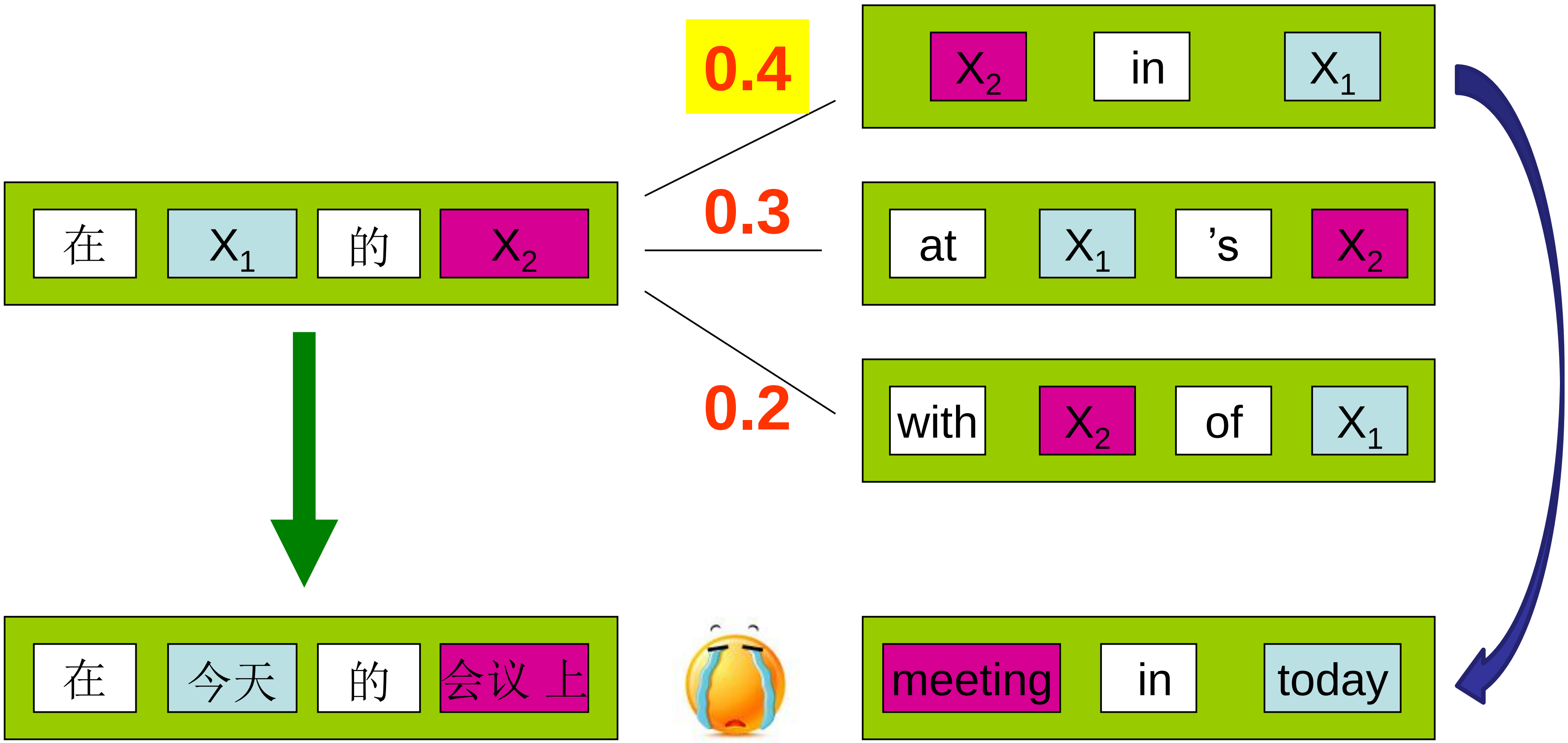
$$X \ \rightarrow < \text{在}\ X_1 \ \text{的}\ X_2,\ X_2 \ \text{in}\ \ X_1 >$$



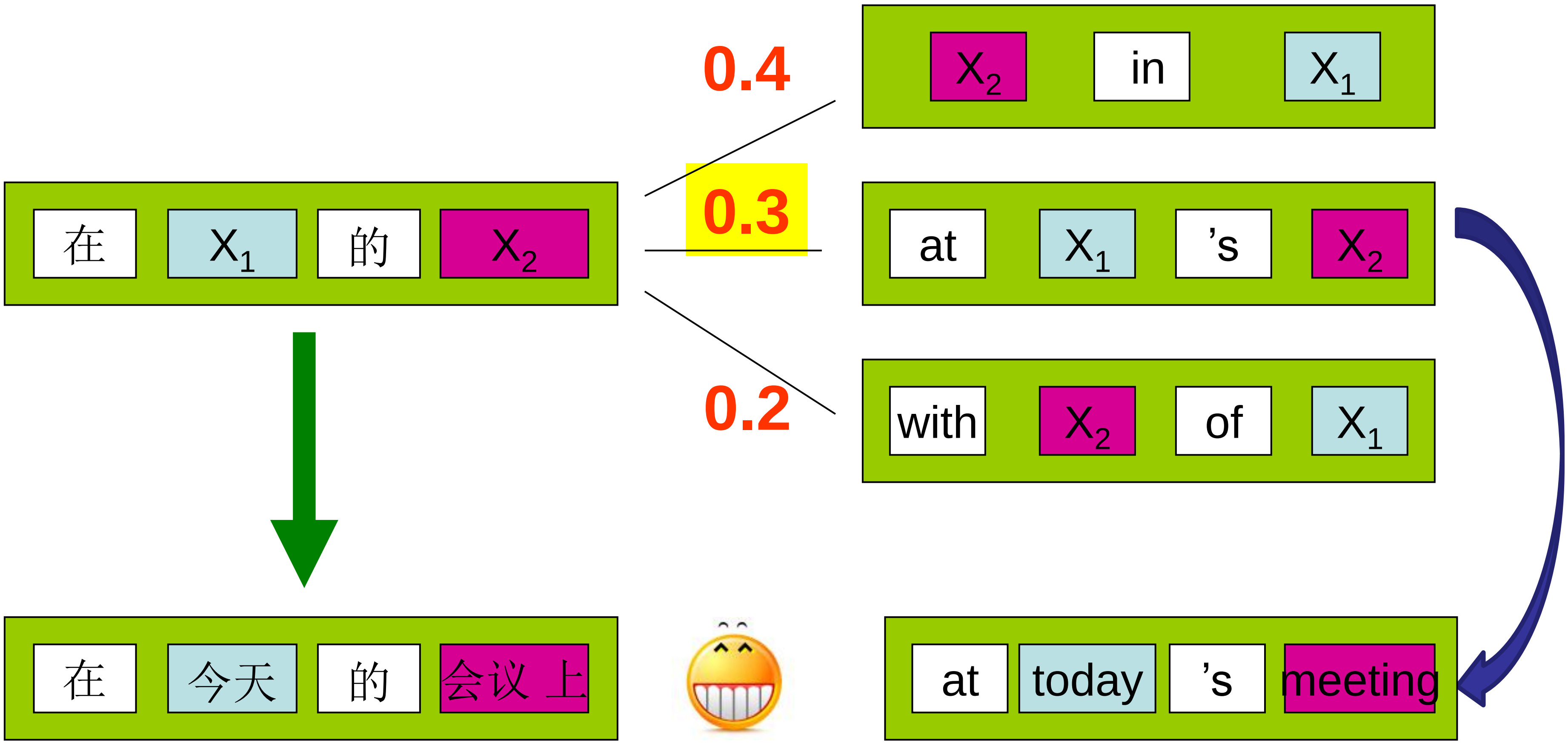
Rule Selection in HPB Model



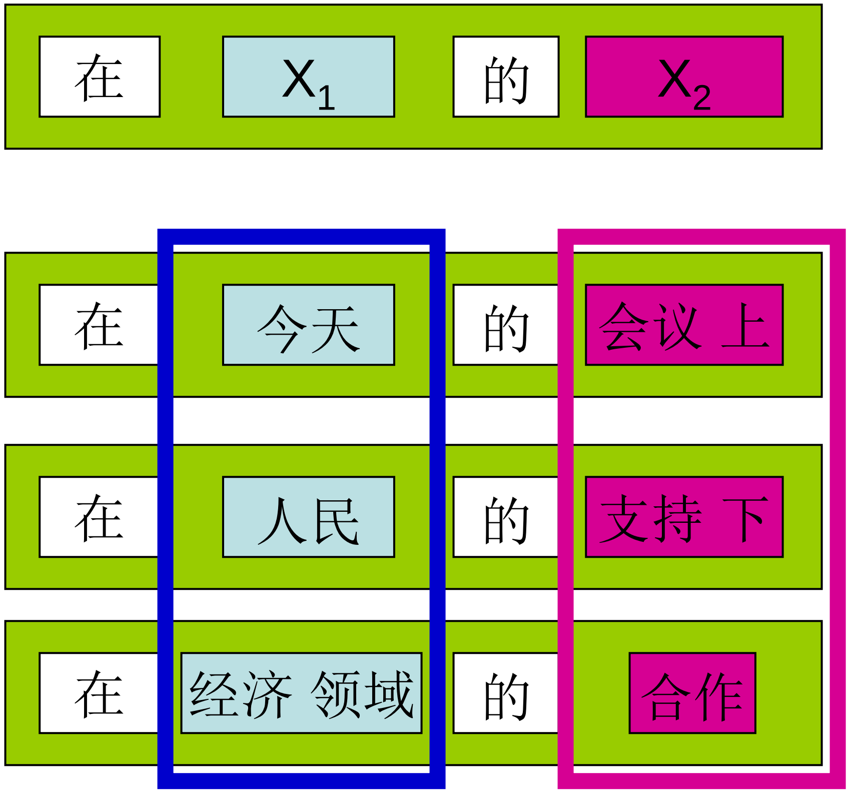
Static Rule Selection



Static Rule Selection



Static Rule Selection



- The corresponding string of X1 and X2 have strong preference for rule selection.
- CARS should be helpful.

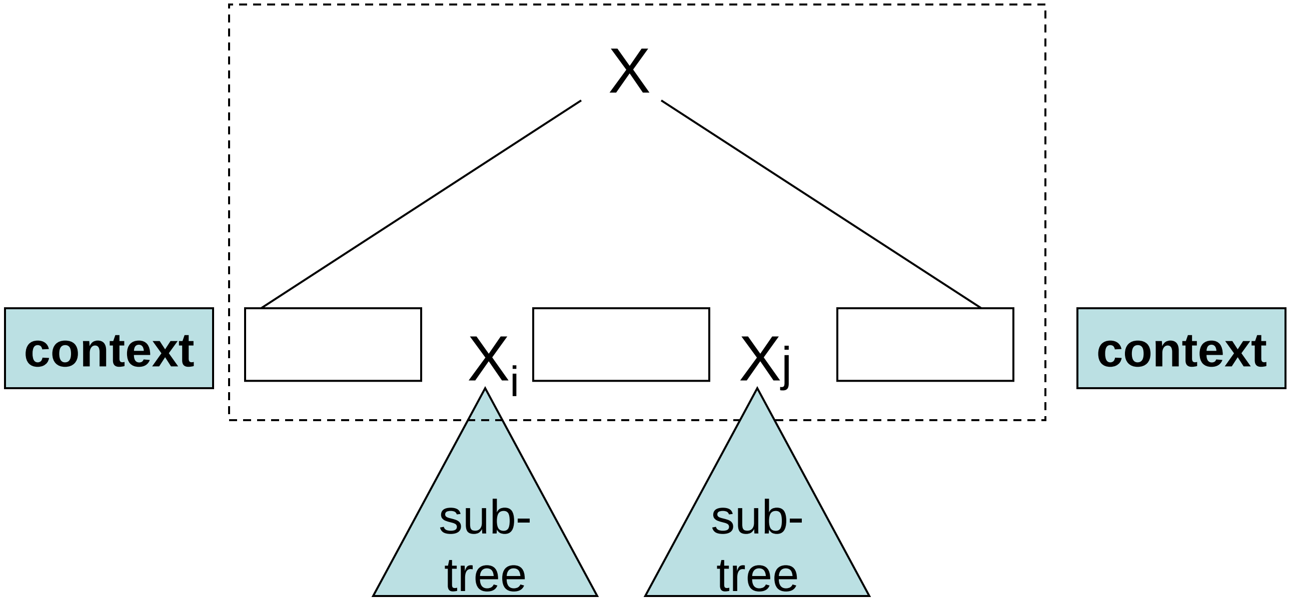
Maximum Entropy RS Model

$$P(r_i|S,C)=\left(\frac{\exp\left(\sum_k\left(h_k\left(\boxed{R,X_1^N}\right)\right)\right)}{\sum_{r_j}\exp\left(\sum_k\left(\lambda_k h_k\left(\boxed{R,X_1^N}\right)\right)\right)}\right)$$

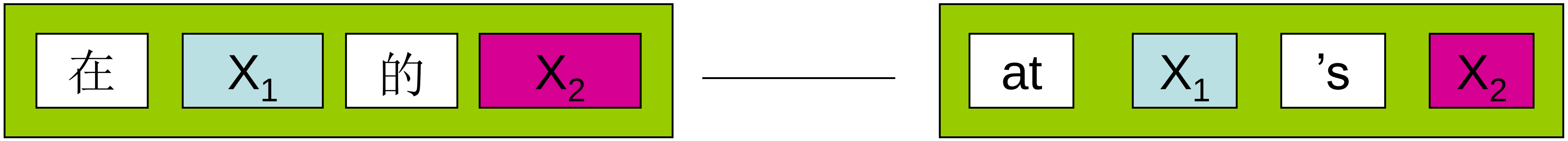
R: Neighbour Context

X_1^N: Variables Context

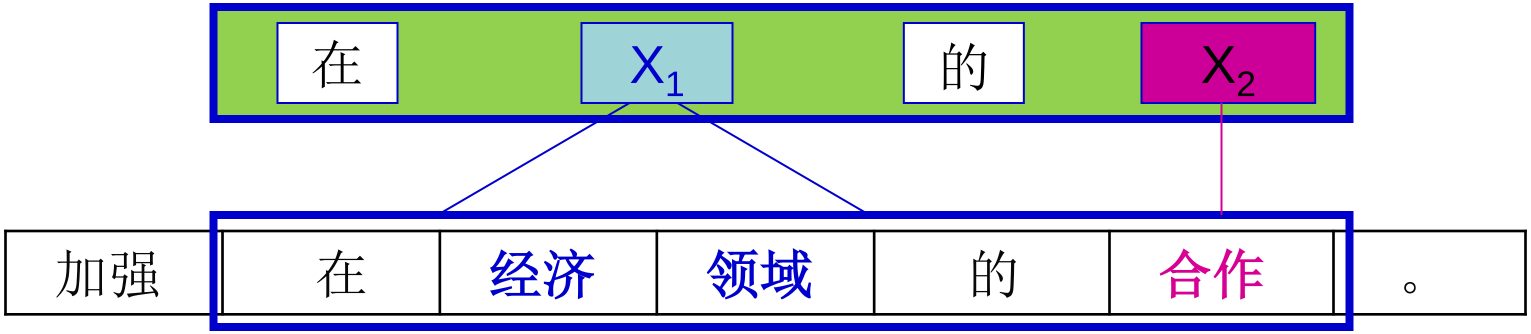
Context for Rule Selection



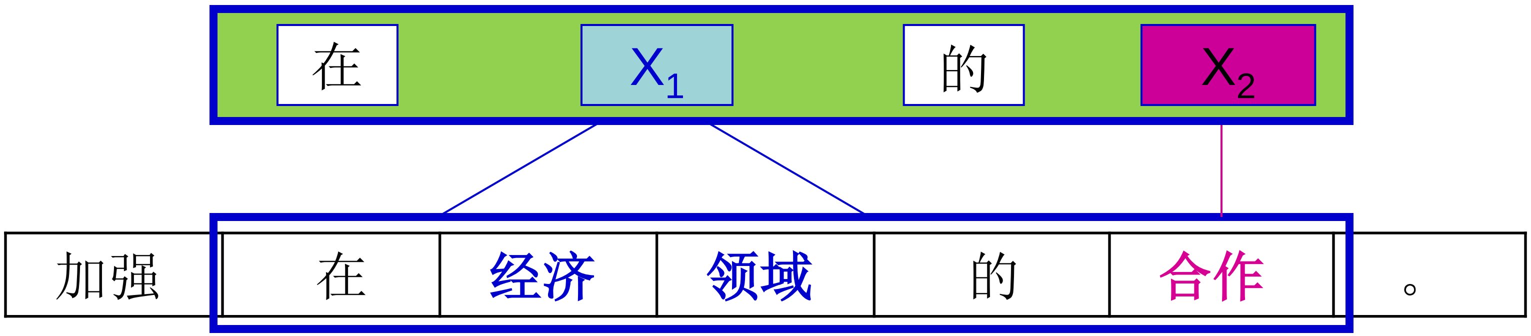
Translation Rule



Source Expression Matching

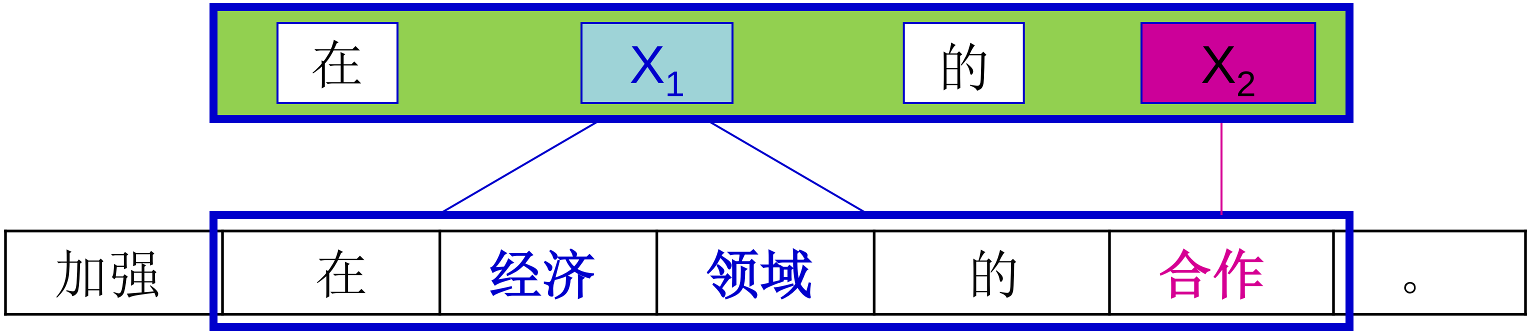


Source Variable Boundary Words



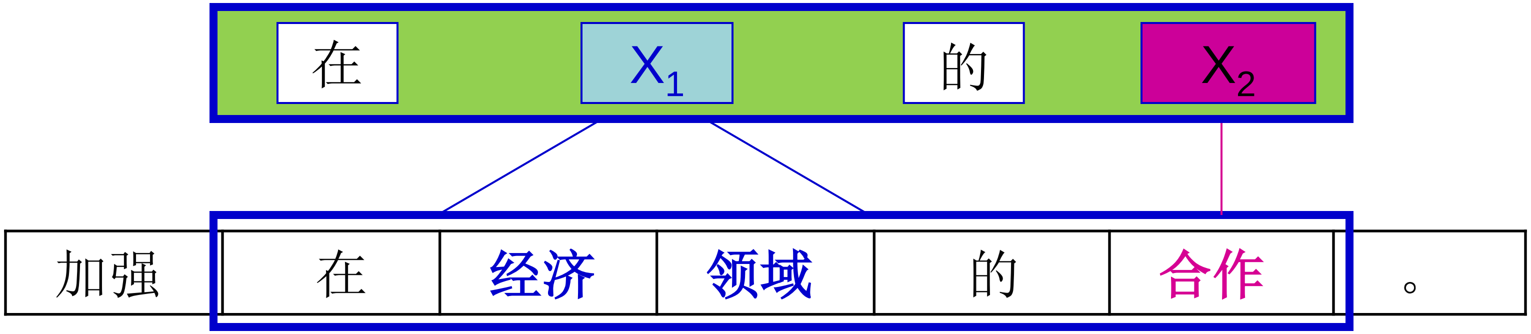
Variable	Feature	Value
X_1	Left Boundary Word	经济
X_1	Right Boundary Word	领域
X_2	Left Boundary Word	合作
X_2	Right Boundary Word	合作

Source Variable Boundary POS



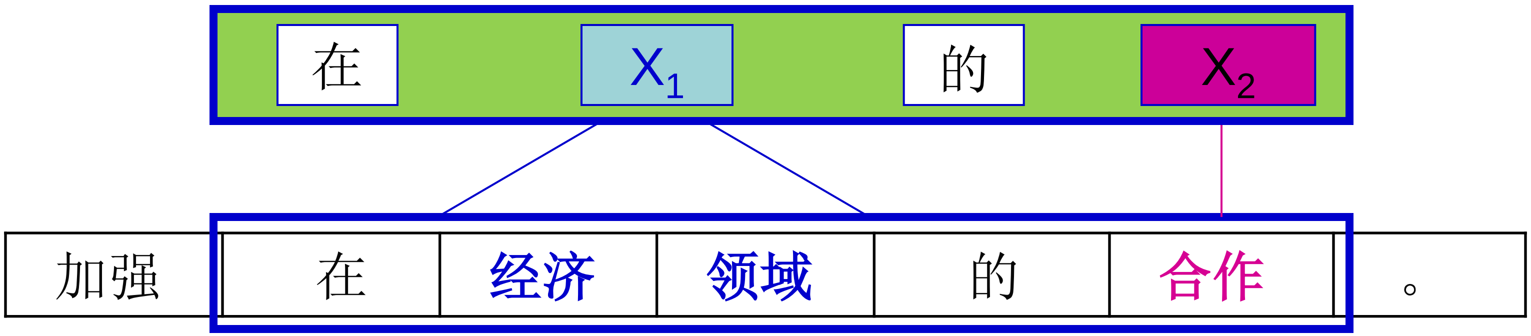
Variable	Feature	Value
X_1	Left Boundary POS	Noun
X_1	Right Boundary POS	Noun
X_2	Left Boundary POS	Noun
X_2	Right Boundary POS	Noun

Source Variable Lengths



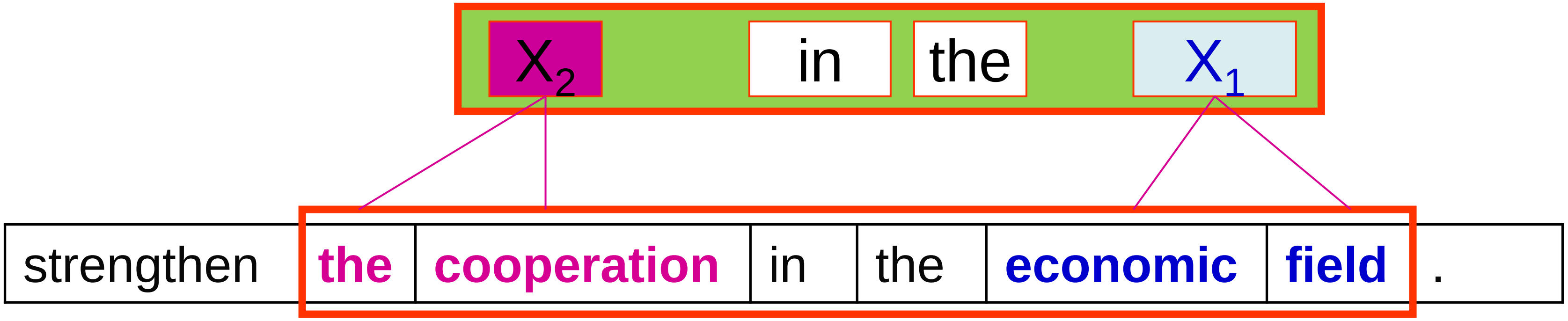
Variable	Feature	Value
X_1	Length	2
X_2	Length	1

Source Neighbour Words and POS

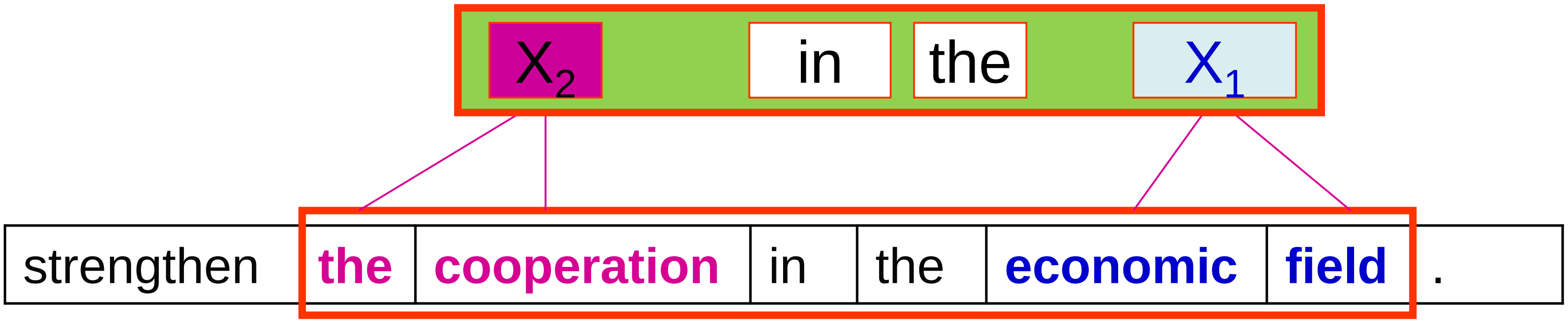


Neighbour	Feature	Value
	Left Word	加强
	Left POS	VERB
	Right Word	。
	Right POS	PUNCT

Target Expression Instantiation

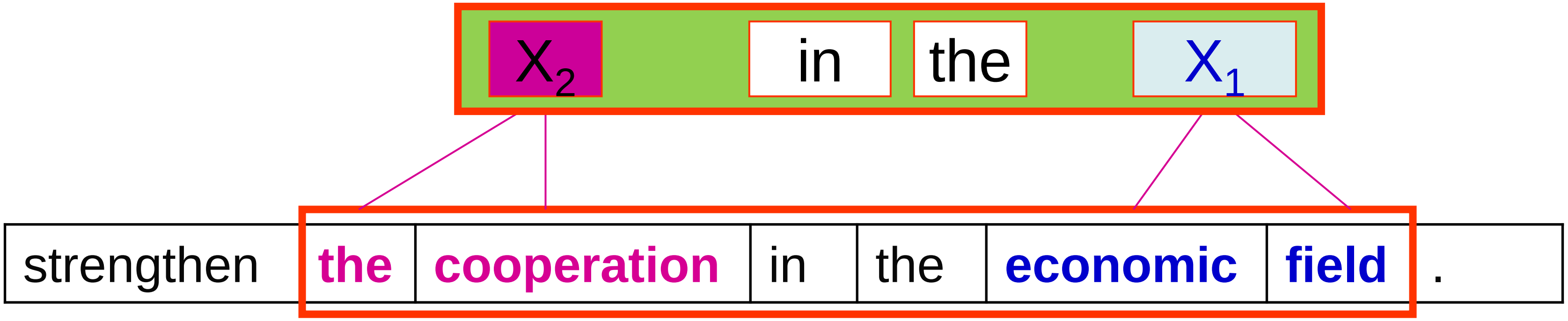


Target Variable Boundary Words



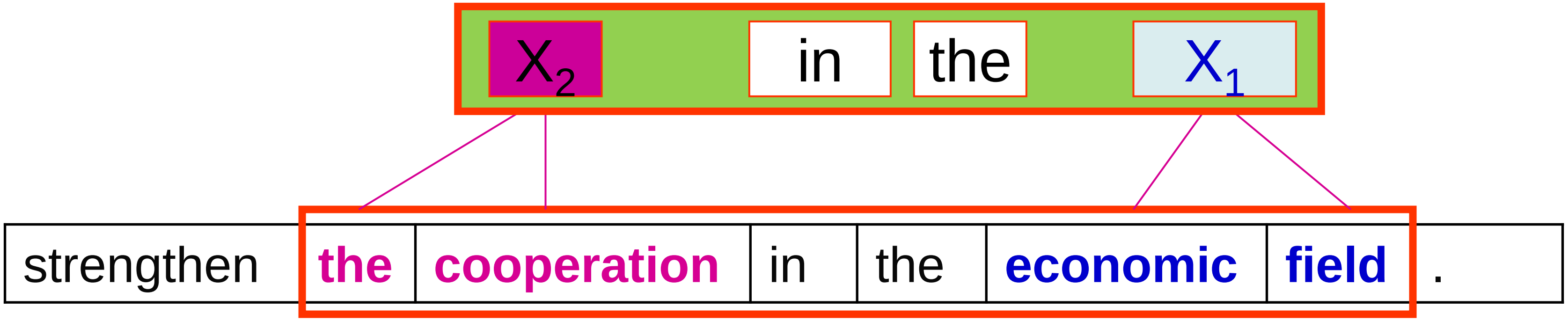
Variable	Feature	Value
X_1	Left Boundary Word	economic
X_1	Right Boundary Word	field
X_2	Left Boundary Word	the
X_2	Right Boundary Word	cooperation

Target Variable Boundary POS



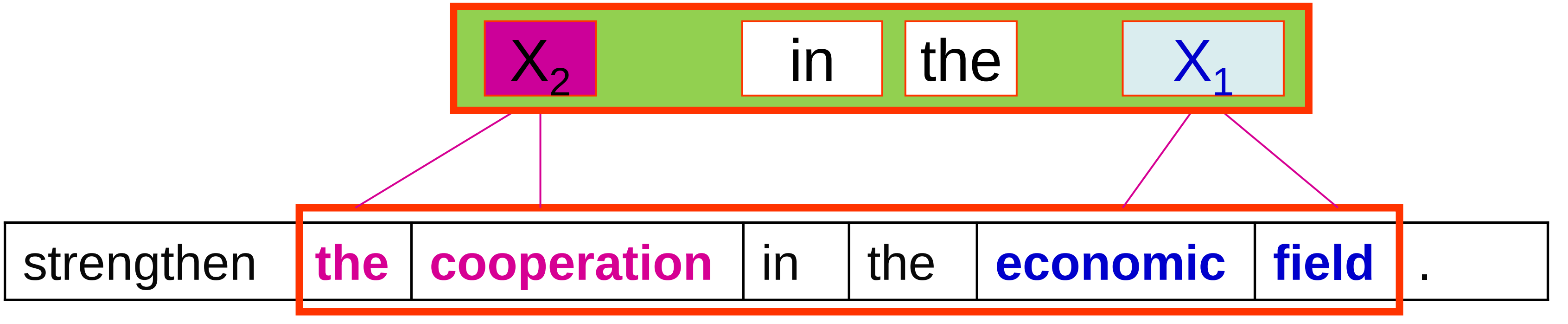
Variable	Feature	Value
X_1	Left Boundary POS	ADJ
X_1	Right Boundary POS	NOUN
X_2	Left Boundary POS	DET
X_2	Right Boundary POS	NOUN

Target Variable Lengths



Variable	Feature	Value
X_1	Length	2
X_2	Length	2

Target Neighbour Words and POS



Inapplicable because we use
a bottom-up decoding manner

Experiment Settings

- Chinese-to-English translation
- Baseline: Reimplementation of Hiero (Chiang 2005)
- Corpus:

Task Name	Training corpus	Dev. set	Test set
IWSLT05	BTEC (40k sent. 354k + 378k)	IWSLT04 (500 sent.)	IWSLT05 (506 sent.)
NIST03	FBIS (239k sent. 6.9M + 8.9M)	NIST02 (878 sent.)	NIST03 (919 sent.)

Experiment Results

System		NIST03 (BLEU-4%)	IWSLT05 (BLEU-4%)
Baseline		28.05	56.20
Baseline +MERS	lexical features (source-side)	28.26	56.51
	POS features	28.78	56.95
	lexical features (source-side) + POS features	28.89	56.99
	lexical features (source-side) + POS features + length features (source-side)	28.96	57.10
	All features (source + target)	**29.02**	**57.20**

* case insensitive

↑ **0.97** ↑ **1.0**

Better Phrase Translation: for **terminal** rules

Source	恐怕　这趟　航班　已经　订　满　了 。
Baseline	$X \rightarrow < X_1$ 订 满, X_1 booked $>$
	I'm afraid already booked for this flight .
Baseline +MERS	$X \rightarrow < X_1$ 订 满, X_1 full $>$
	I'm afraid this flight is full .

Better Phrase Reordering: for **nonterminal** rules

source	… 联合国 安全 理事会 的 五 个 常任 理事国…
Baseline	$X \to < X_1$ 的 X_2, the $X_1\ X_2 >$
	… the United Nations Security Council five permanent members …
Baseline +MERS	$X \to < X_1$ 的 X_2, X_2 of $X_1 >$
	… the five permanent members of the UN Security Council …

Summary

- A MERS model was proposed for hierarchical phrase-based model
- Features used in MERS model:
  - Boundary words and POS tags of internal variables
  - Boundary words and POS tags of neighbours
- MERS help to improve the system performance significantly

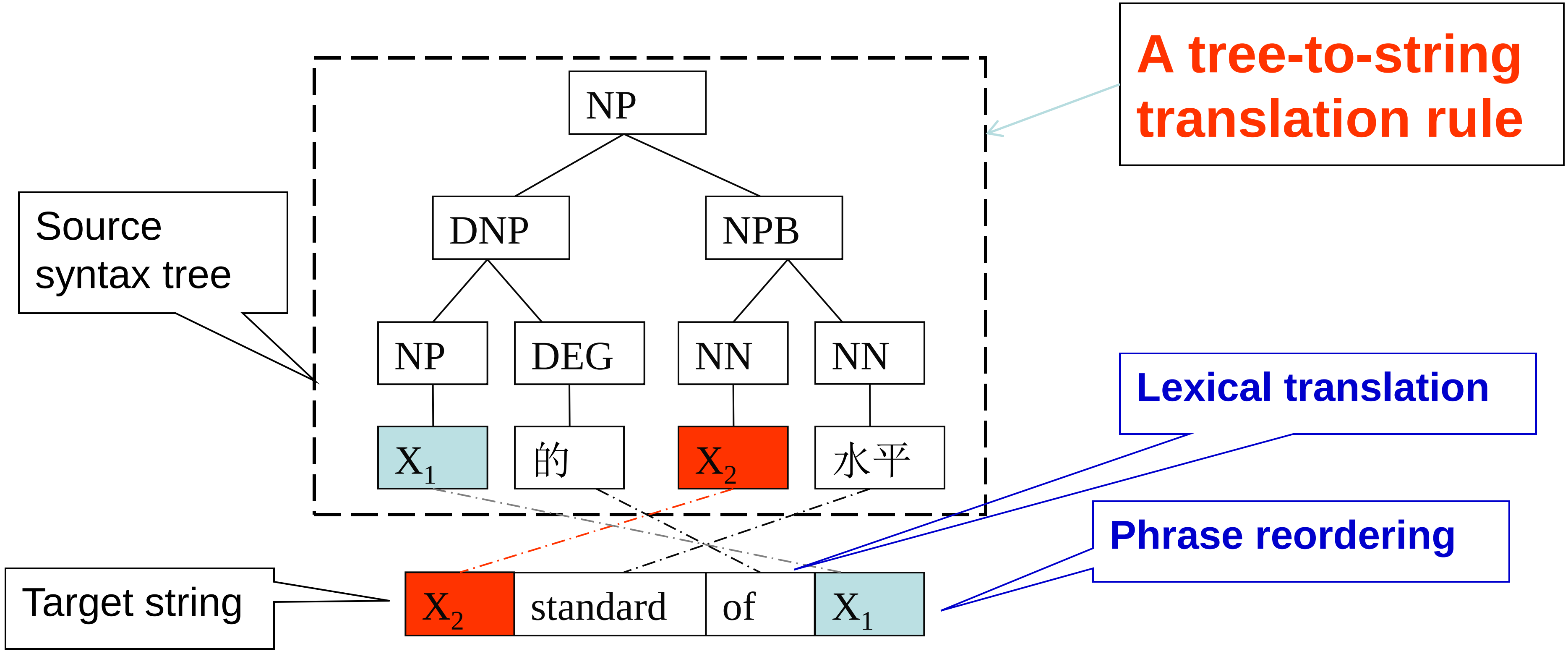
CARS Application Examples

- CARS for Bracketing Transduction Grammar
- CARS for Hierarchical Phrase-based Model
- CARS for Tree-to-String Model
- CARS using Topic Model
- CARS for Agglutinative Language Translation

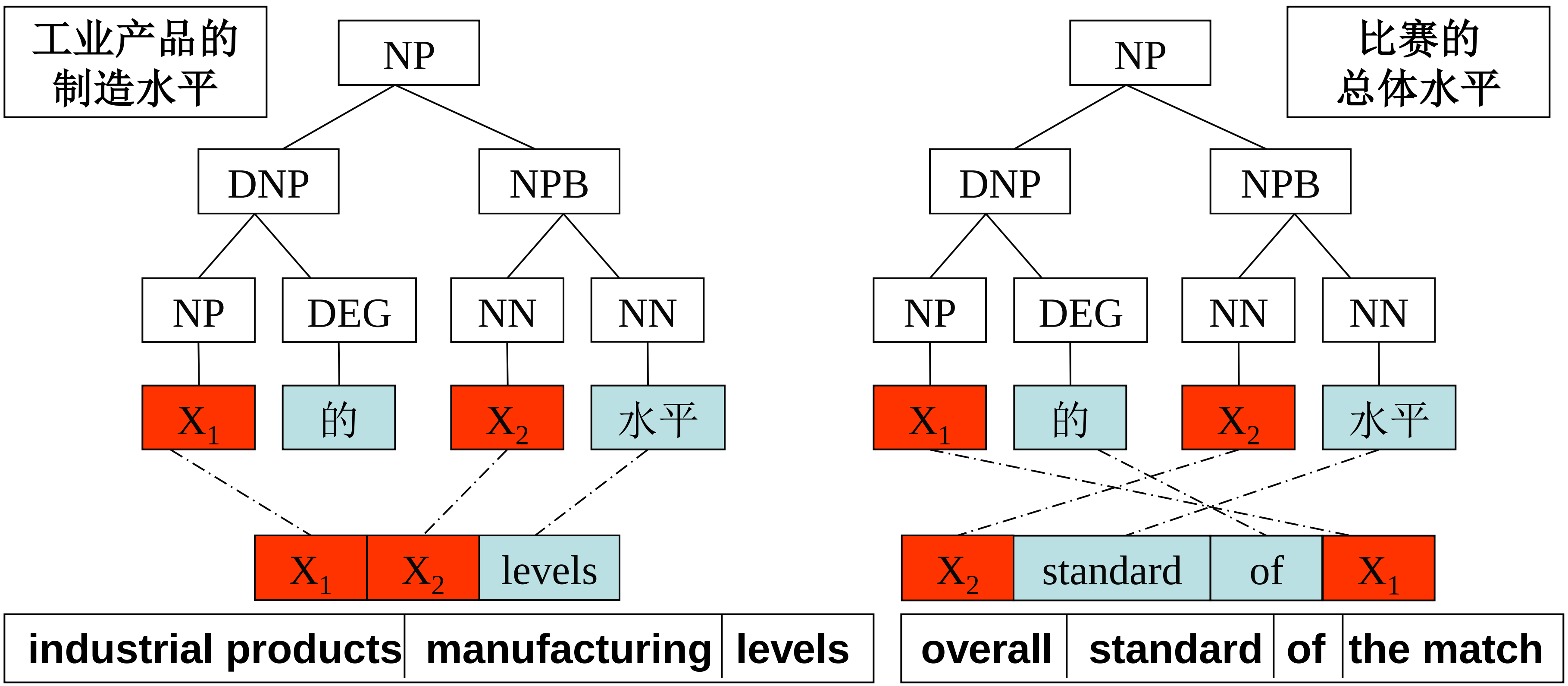
Tree-to-String Model

Yang Liu et al. ACL2006

Liang Huang et al. AMTA2006



Rule Selection Problem



Maximum Entropy RS Model

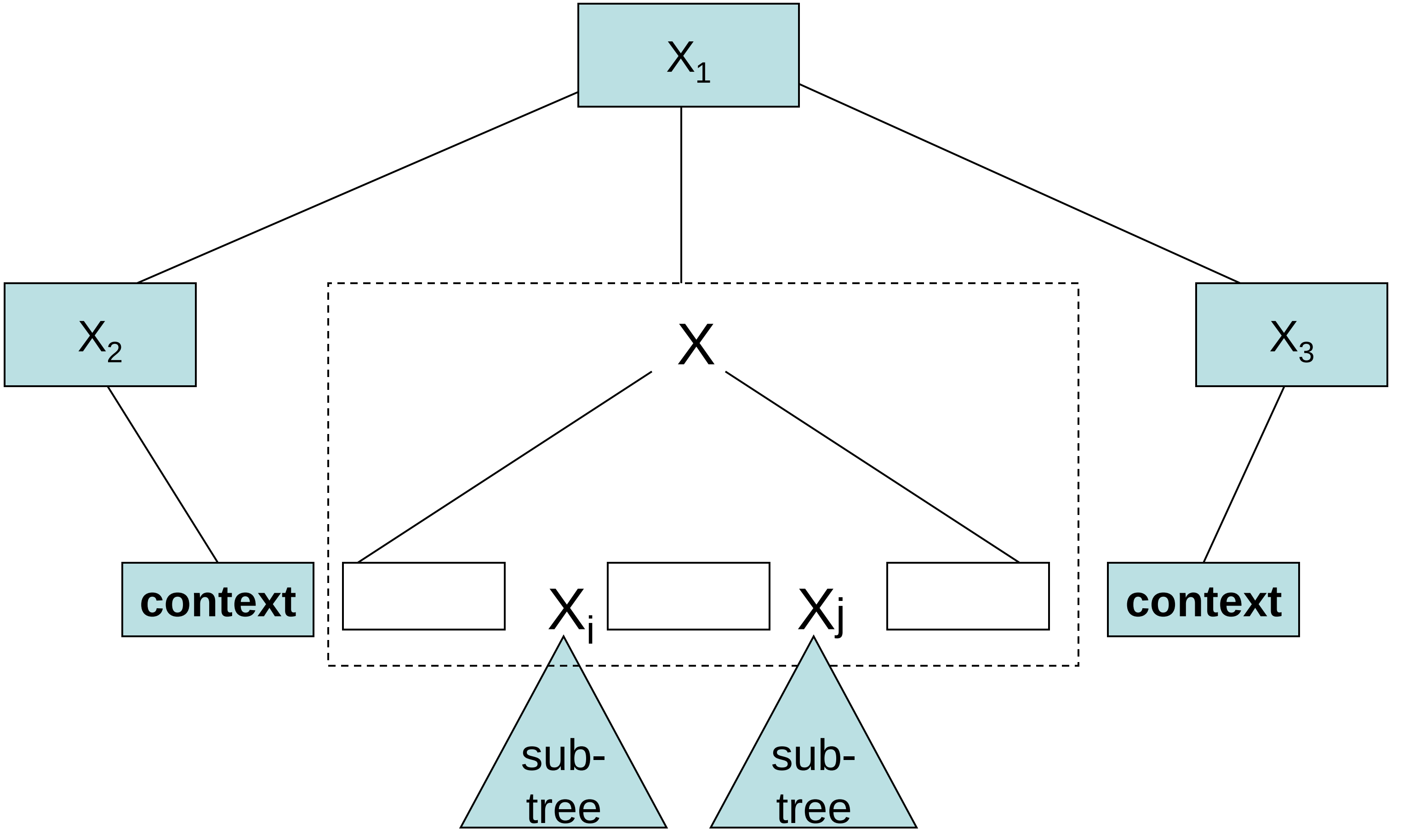
$$P(r_i|S,C)=\left(\frac{\exp(\sum_k(h_k(\boxed{R,Y,X_1^N})))}{\sum_{r_j}\exp(\sum_k(\lambda_k h_k(\boxed{R,Y,X_1^N})))}\right)$$

R: Neighbours

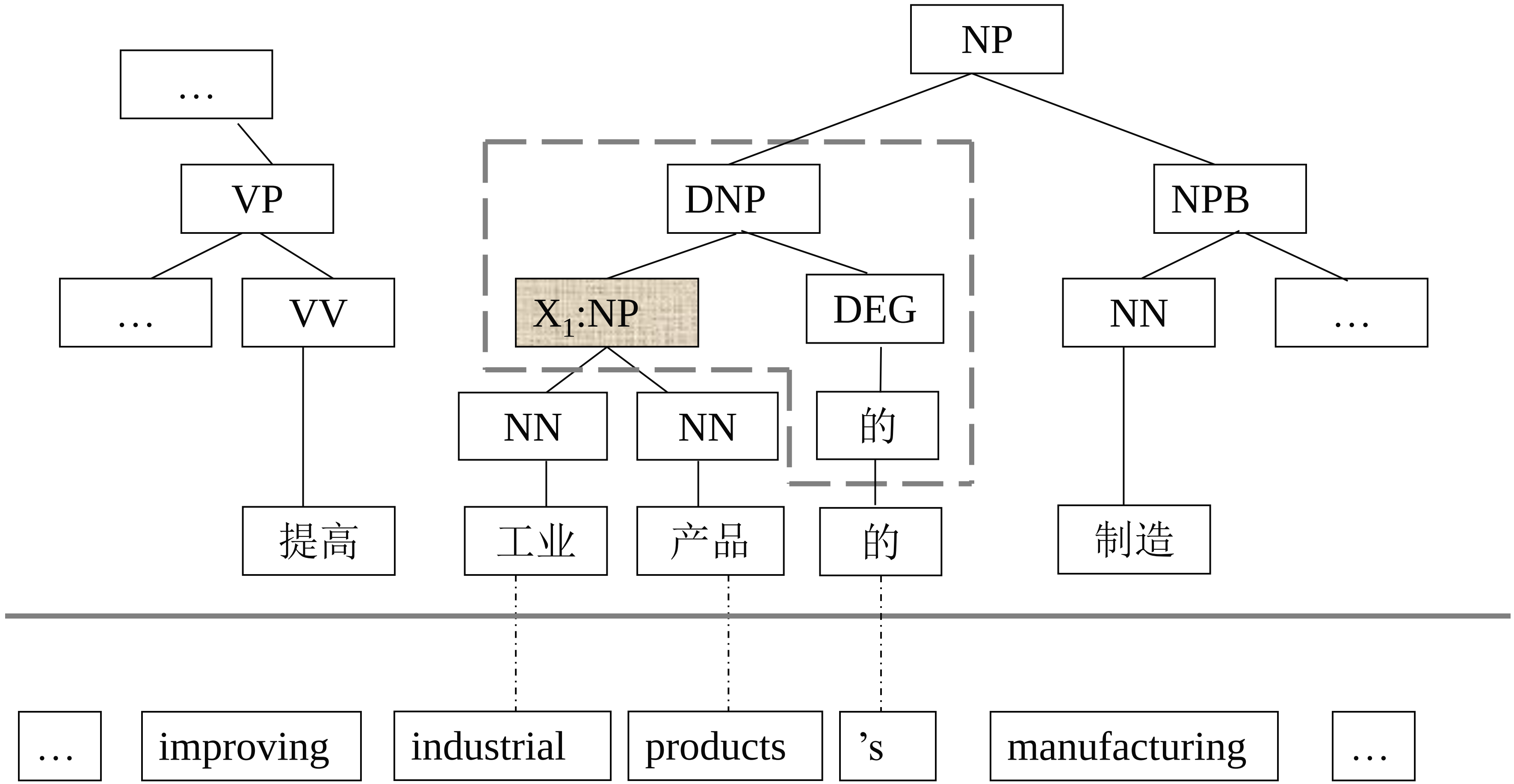
Y: Syntax Tree Context

X_1^N: Internal Variables in Rules

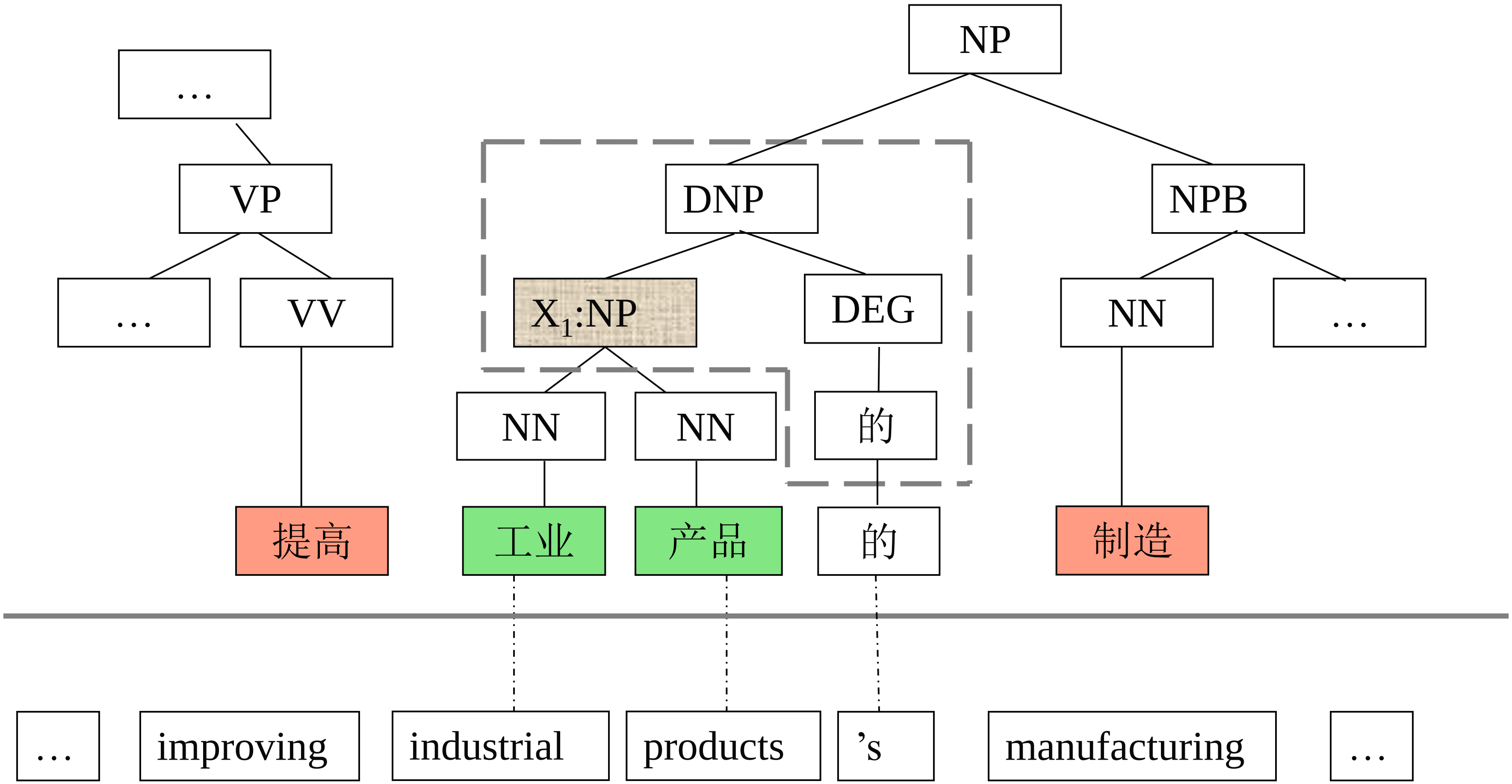
Context for Rule Selection



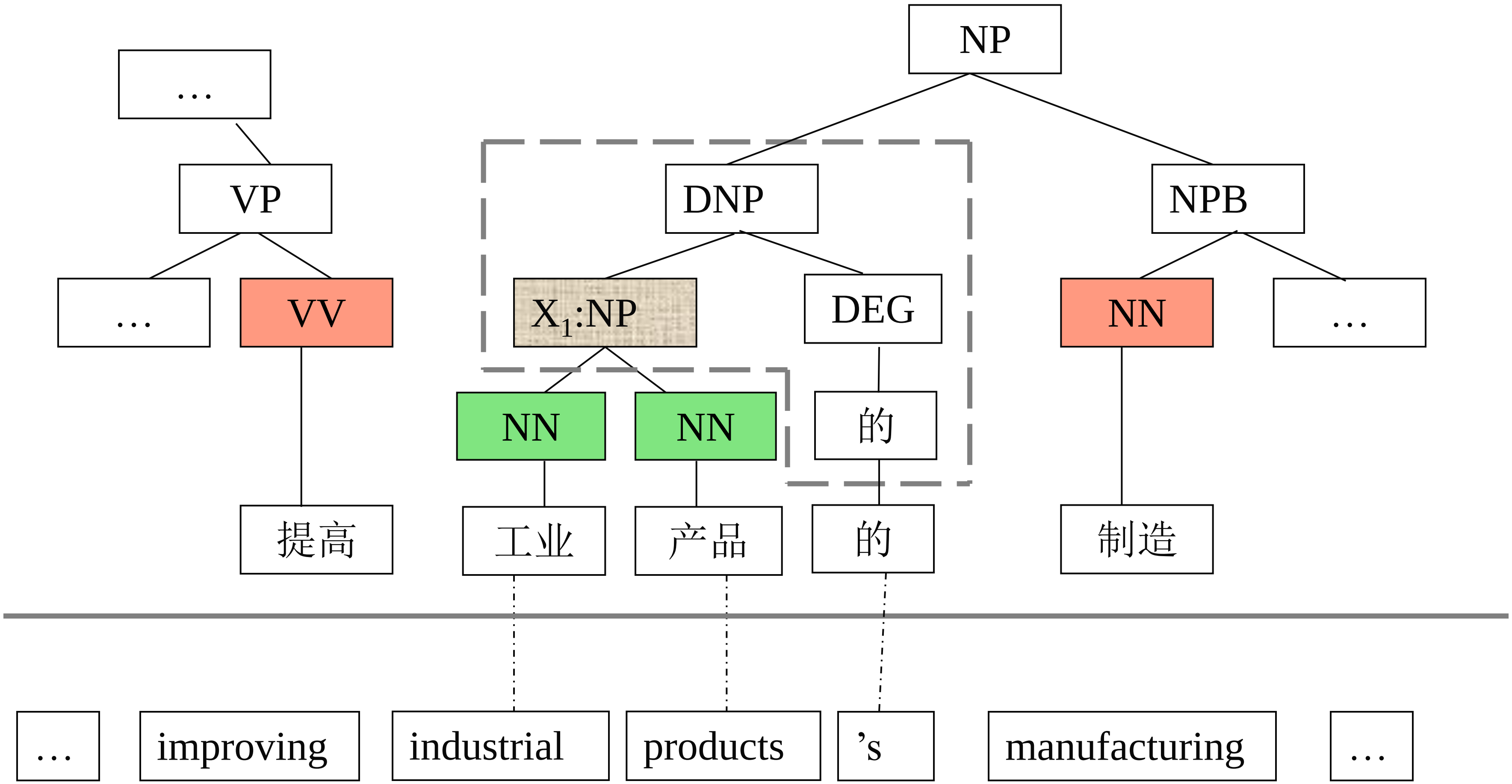
Feature Definition



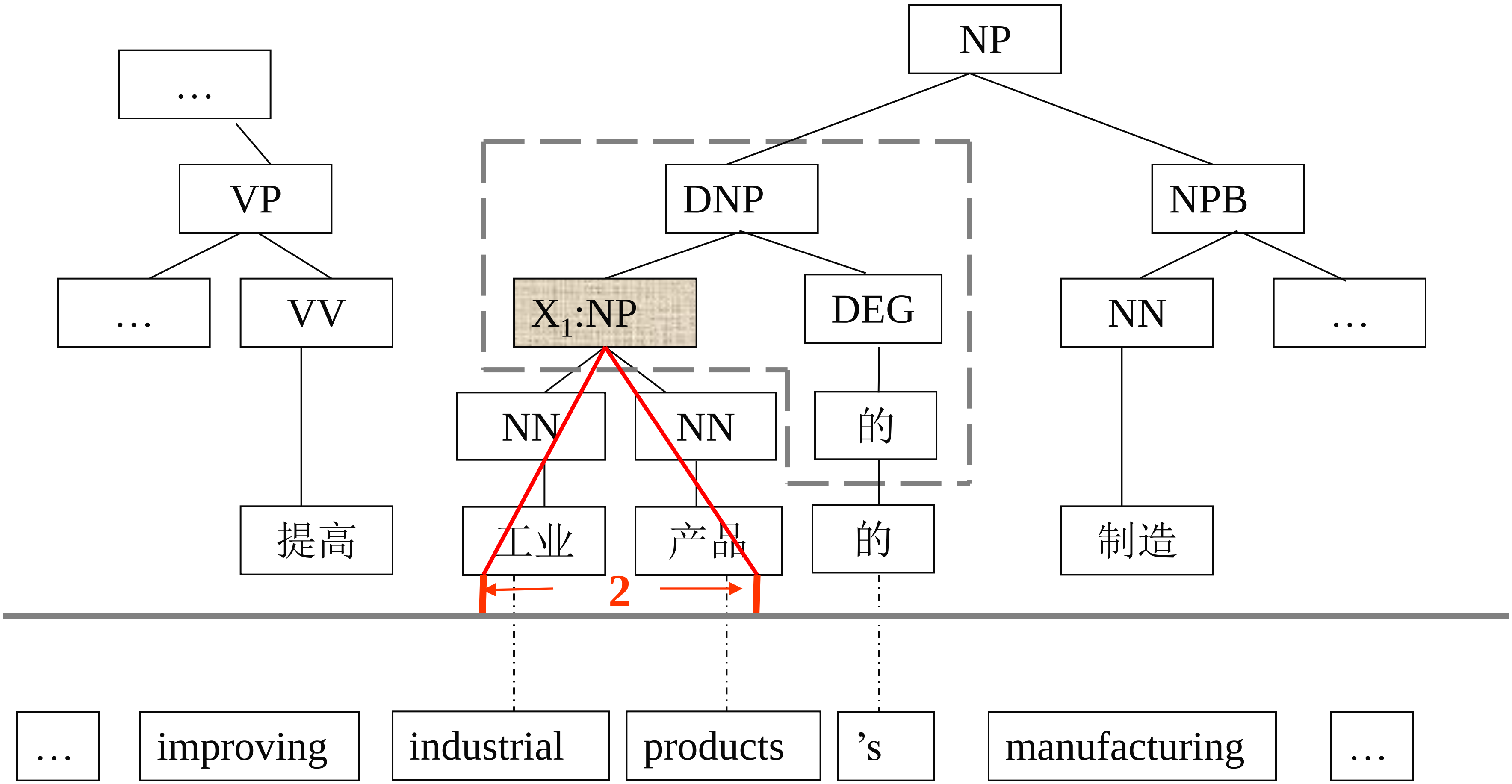
Feature Definition: Lexical Features (LF)



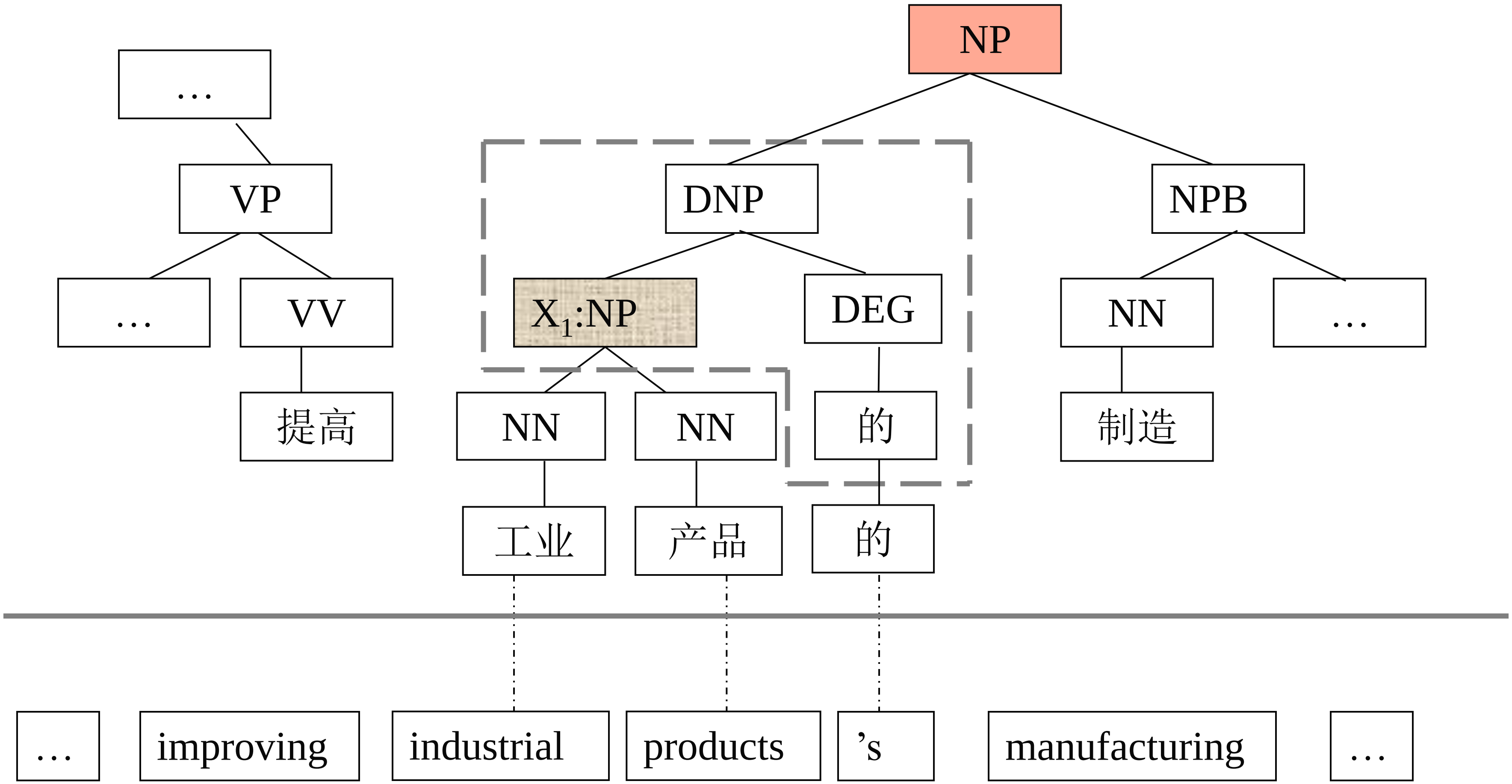
Feature Definition: POS Features (POSF)



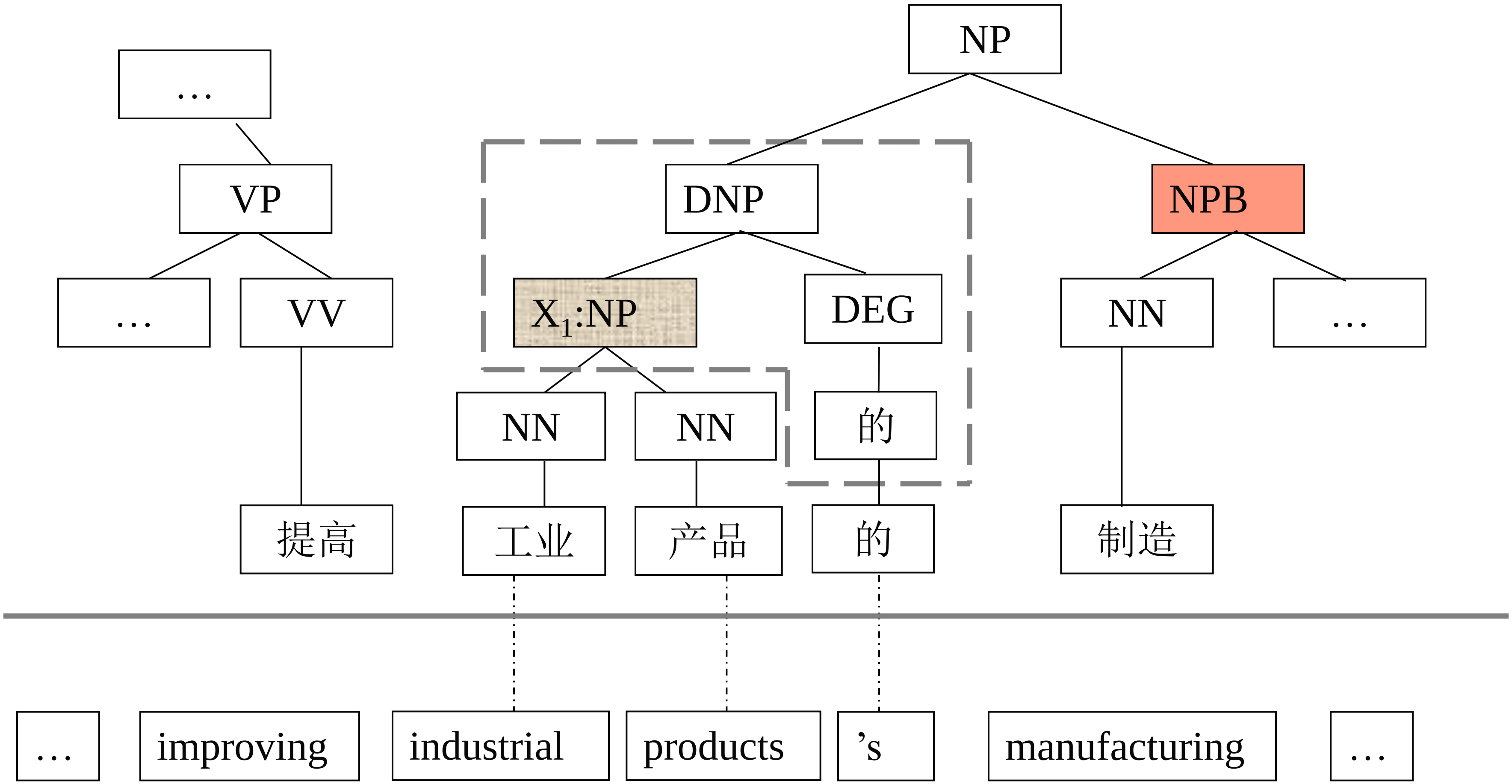
Feature Definition: Span Features (SPF)



Feature Definition: Parent Feature (PF)



Feature Definition: Sibling Features (SBF)



Experiments

- Chinese-to-English translation
- Baseline: Lynx (Liu Yang, et al., 2006), the state-of-the-art syntax-based SMT system
- Corpus:

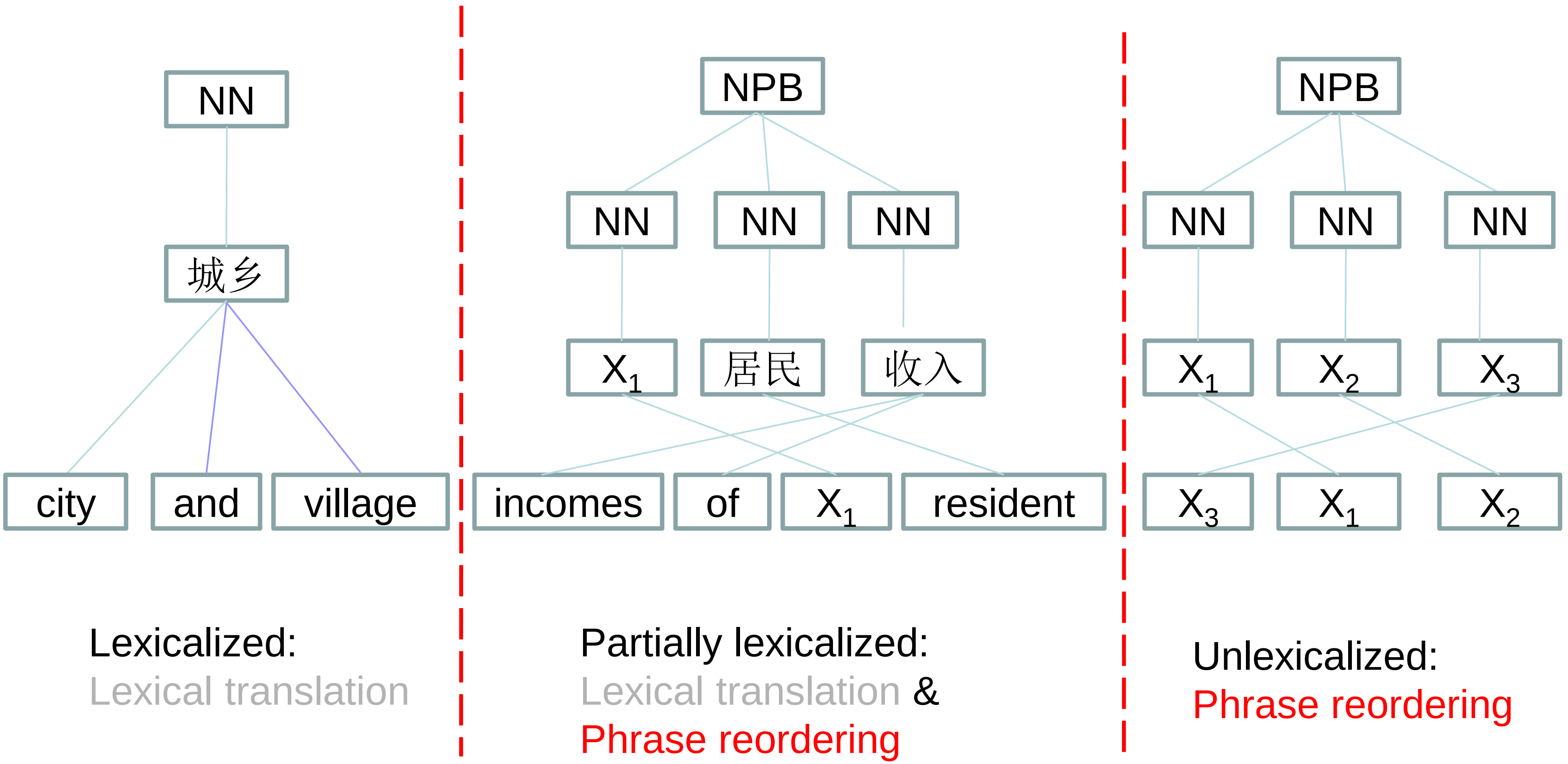
<table>
<tr><th>Training corpus</th><th>Dev. set</th><th>Test set</th></tr>
<tr><td rowspan="2">FBIS (239k sent.
6.9M + 8.9M)</td><td rowspan="2">NIST02
(878 sent.)</td><td>NIST03
(919 sent.)</td></tr>
<tr><td>NIST05
(1082 sent.)</td></tr>
</table>

Results

System		NIST03 (BLEU-4%)	NIST05 (BLEU-4%)
Lynx		**26.15**	**26.09**
Lynx +MERS	LF	26.12	26.32
	POSF	26.36	26.21
	PF	26.17	25.90
	SBF	26.47	26.08
	LF+POSF	26.61	26.59
	LF+POSF+SPF	26.70	26.44
	LF+POSF+PF	26.81	26.56
	LF+POSF+SBF	26.68	26.89
	ALL	**27.05**	**27.28**

↑ **0.9** ↑ **1.19**

Three kinds of TATs

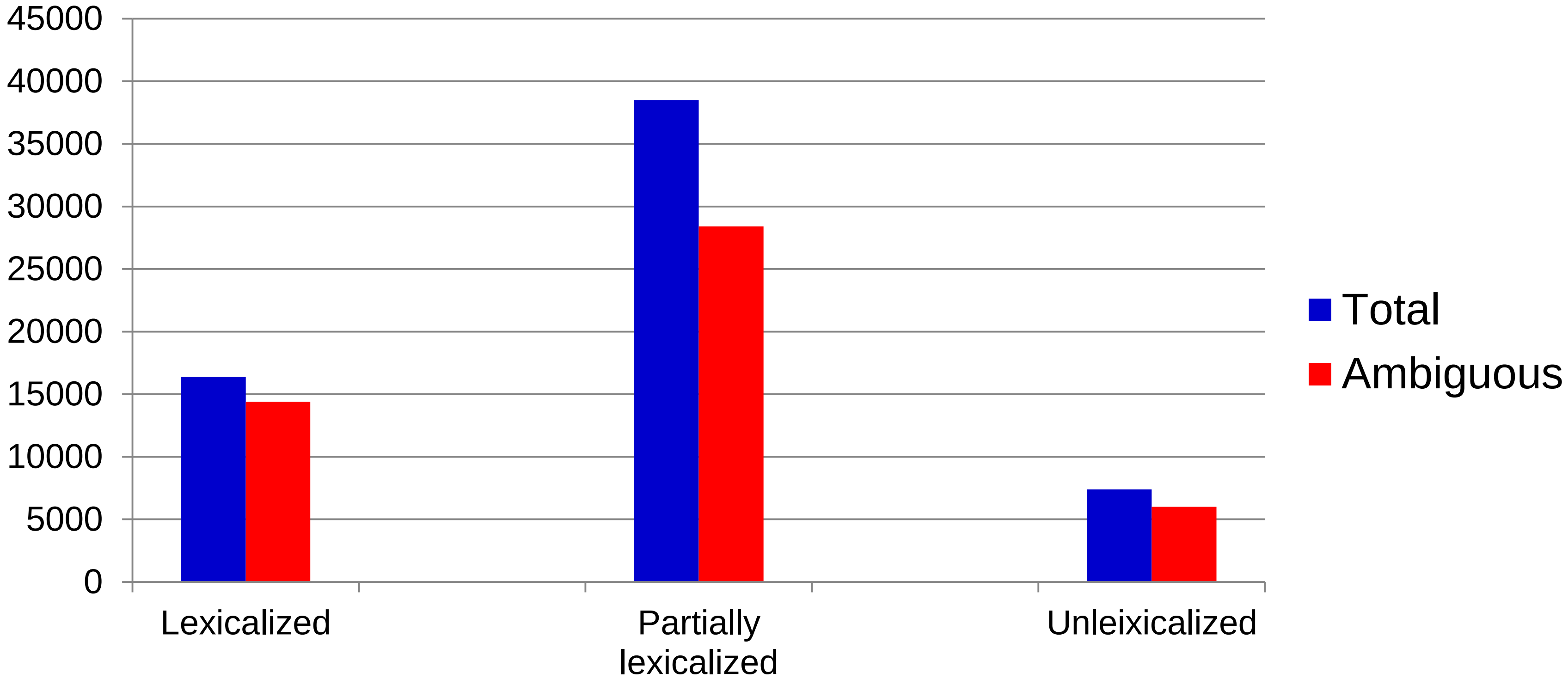


Lexicalized:
Lexical translation

Partially lexicalized:
Lexical translation &
Phrase reordering

Unlexicalized:
Phrase reordering

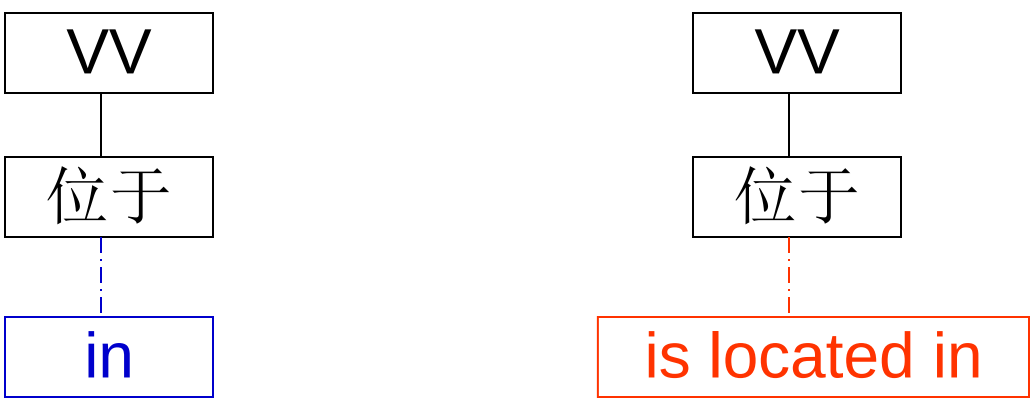
Statistical Info. of source trees for Test Sets



More than 78% source trees are ambiguous!

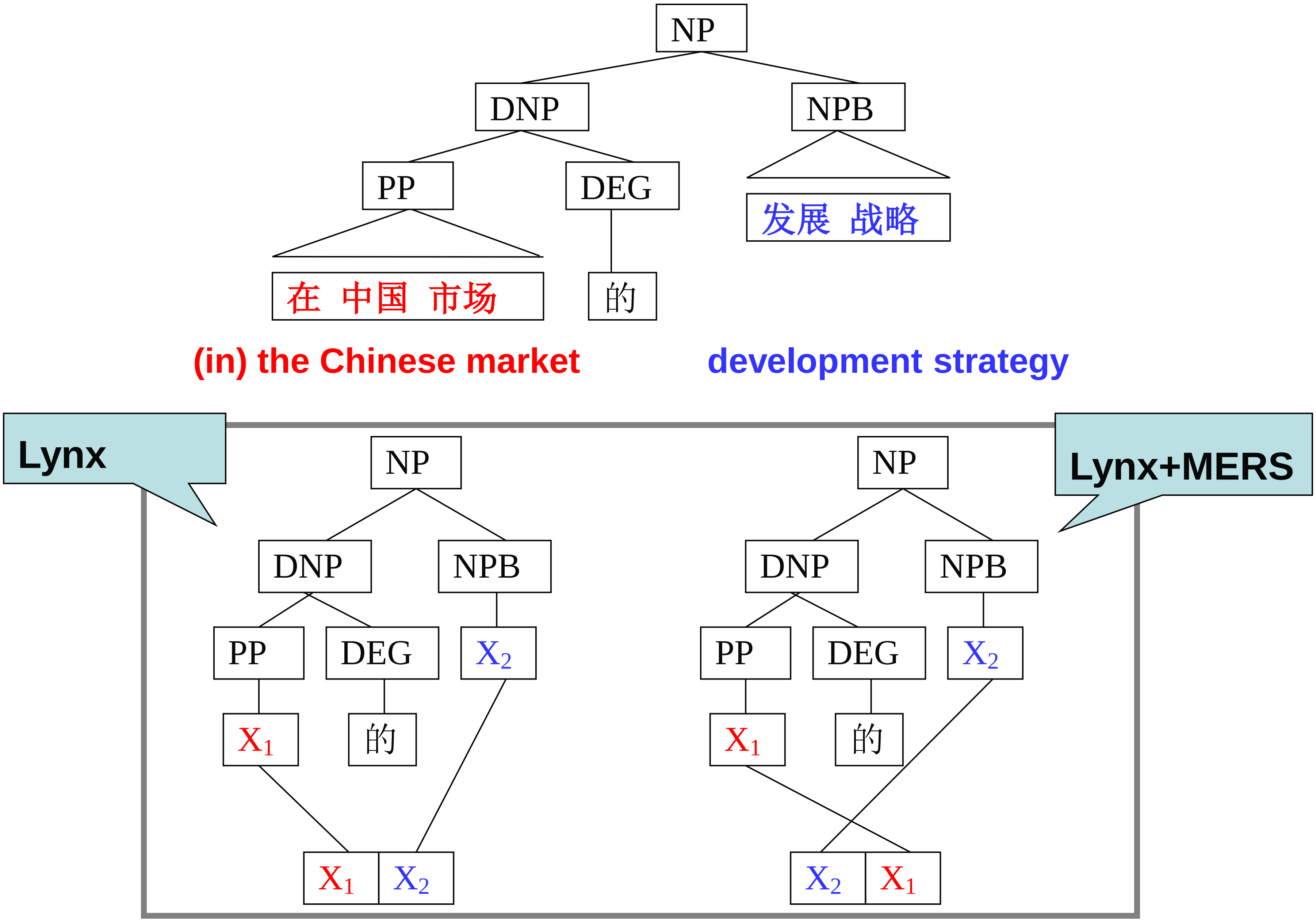
Better Lexical Translation

Source	马耳他 位于 欧洲 南部
Lynx	Malta in southern Europe
Lynx+MERS	Malta is located in southern Europe



Better Phrase Reordering

Source	按照 在 中国 市场 的 发展 战略 , …
Lynx	Accordance with the Chinese market development strategy , …
Lynx+MERS	According to the development strategy in the Chinese market , …


NP
DNP
NPB
PP
DEG
发展 战略
在 中国 市场
的
(in) the Chinese market
development strategy
Lynx
Lynx+MERS
NP
DNP
NPB
PP
DEG
X2
X1
的
X1 X2
NP
DNP
NPB
PP
DEG
X2
X1
的
X2 X1

Summary

- A MERS model was proposed for tree-to string model
- Features used in MERS model:
  - Boundary words and POS tags of internal variables
  - Boundary words and POS tags of neighbours
  - Syntax labels of parent node and sibling node
- MERS help to improve the system performance significantly

CARS Application Examples

- CARS for Bracketing Transduction Grammar
- CARS for Hierarchical Phrase-based Model
- CARS for Tree-to-String Model
- CARS using Topic Model
- CARS for Agglutinative Language Translation

Rule Selection by Topic

Bank

Mouse

Rule Selection by Topic

银行 Finance

Bank

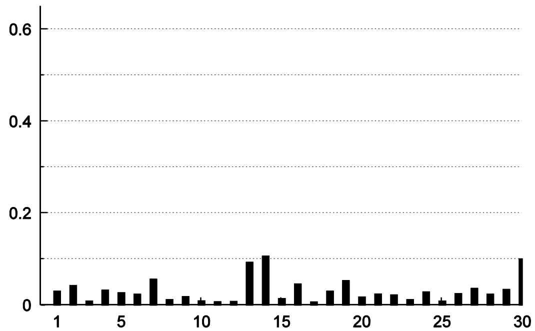
河岸 Geography

老鼠 Biology

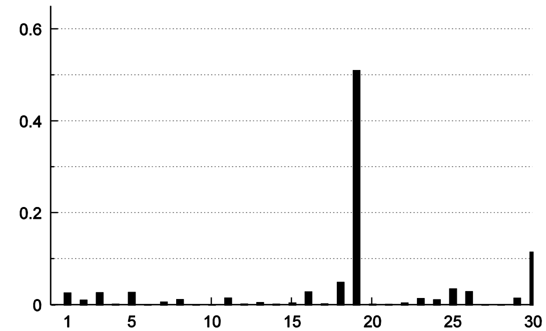
Mouse

鼠标 Computer

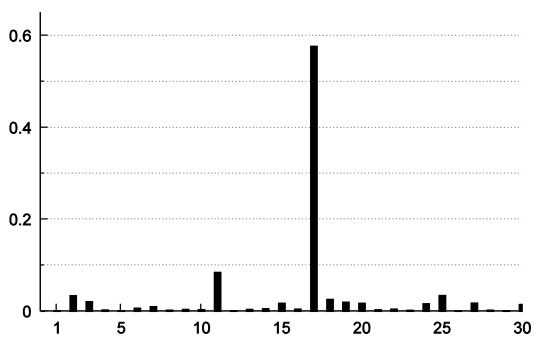
Topic Distribution of Rules



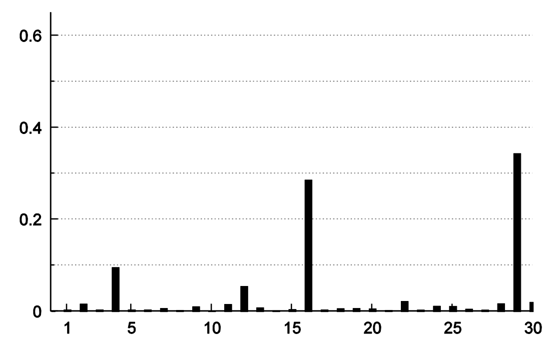
给予 X_1 ⇨ give X_1



给予 X_1 ⇨ grants X_1



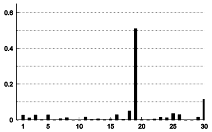
作战 能力 ⇨ operational capacity

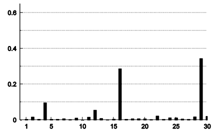


X_1 举行 会谈 X_2 ⇨ held talks X_1 X_2

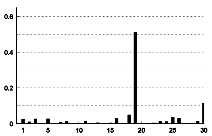
Topic Similarity and Sensitivity

- Topic Similarity Model



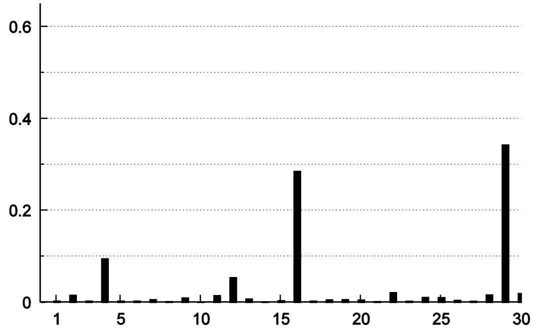


  – Describe the relatedness of rules to topics of given documents
- Topic Sensitivity Model



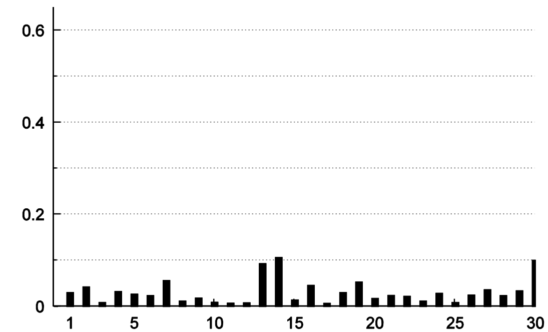
  – Distinguish topic-insensitive rules and topic-sensitive rules

Topic Similarity Model

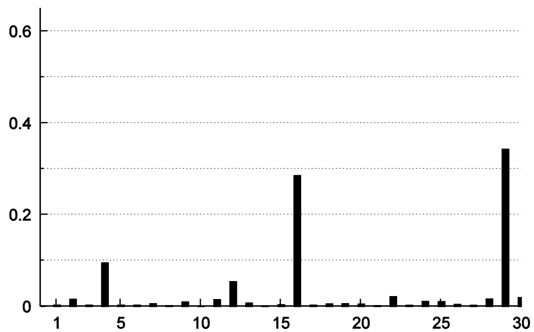


Source Document

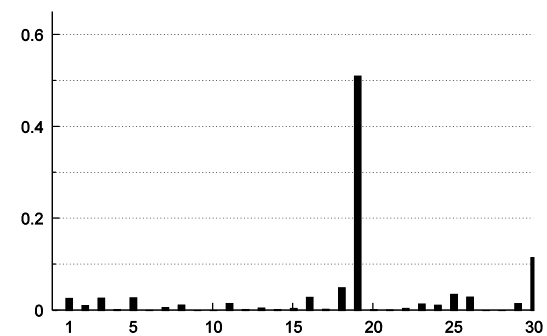
Topic Similarity Model



给予 X_1 ⇨ give X_1

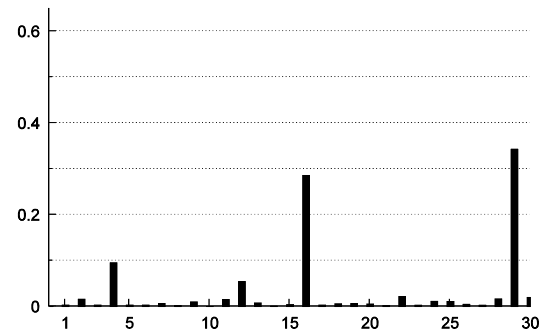


Source Document



给予 X_1 ⇨ grants X_1

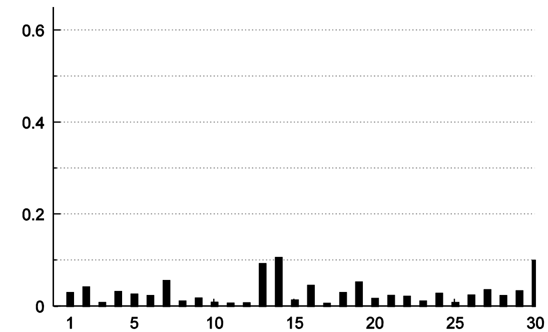
Topic Similarity Model



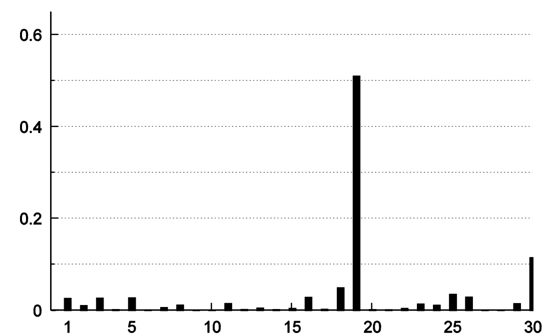
Source Document

Distribution

Distance



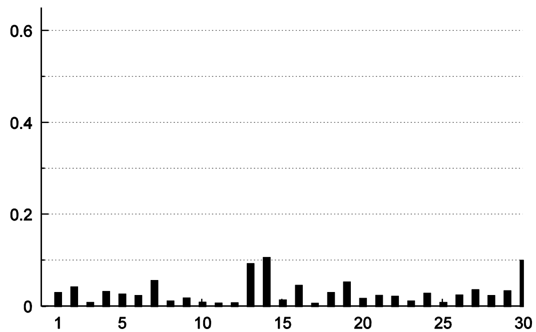
给予 X_1 ⇨ give X_1



给予 X_1 ⇨ grants X_1

$$\sum_{k=1}^{K}\left(\sqrt{\vec{p}(z=k|d)}-\sqrt{\vec{p}(z=k|r)}\right)^2$$

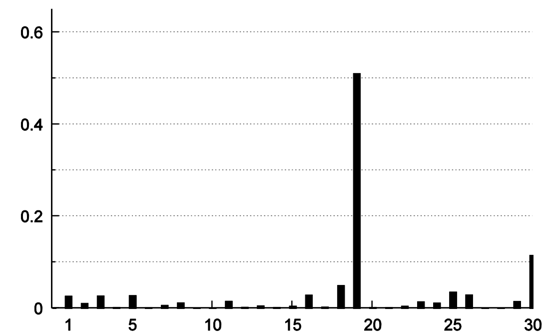
Topic Sensitivity Model



给予 X_1 ⇨ give X_1

Topic-insensitive Rule

Applied in many topics



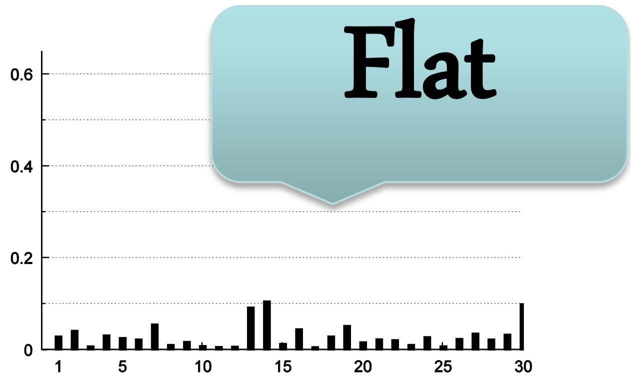
给予 X_1 ⇨ grants X_1

Topic-sensitive Rule

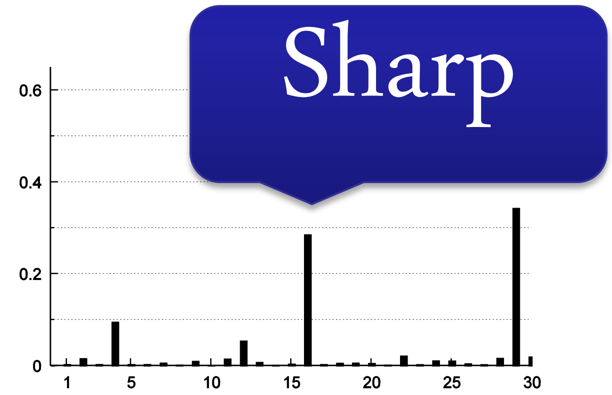
Applied in few topics

- Describe by Entropy as a metric

Topic Sensitivity Model



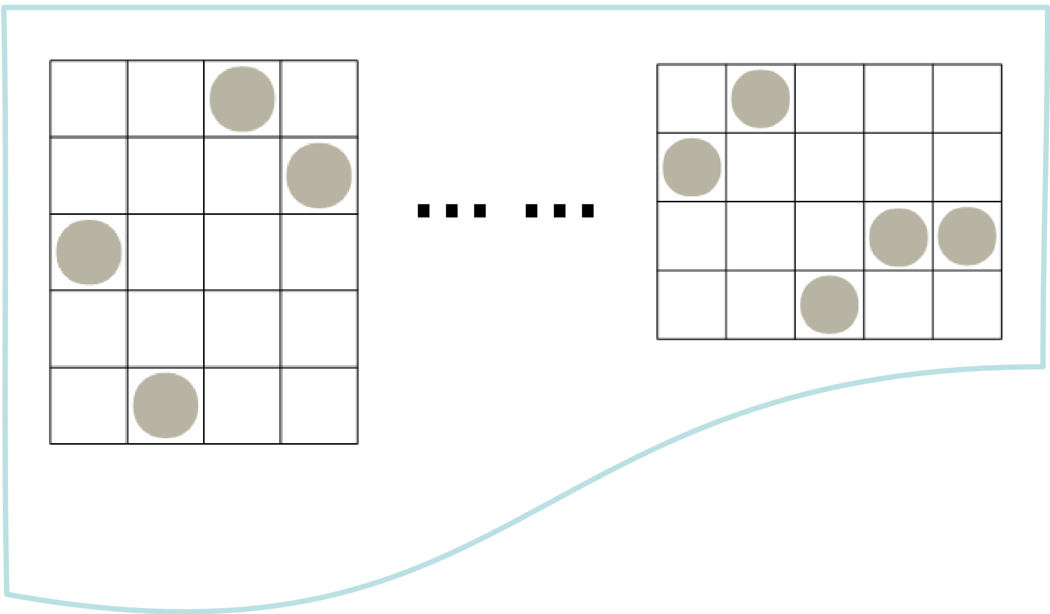
Topic-insensitive Rule

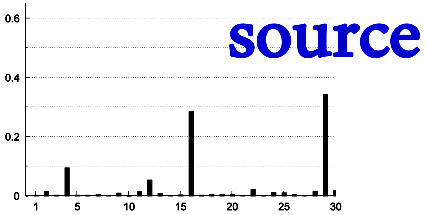


Topic-sensitive Rule

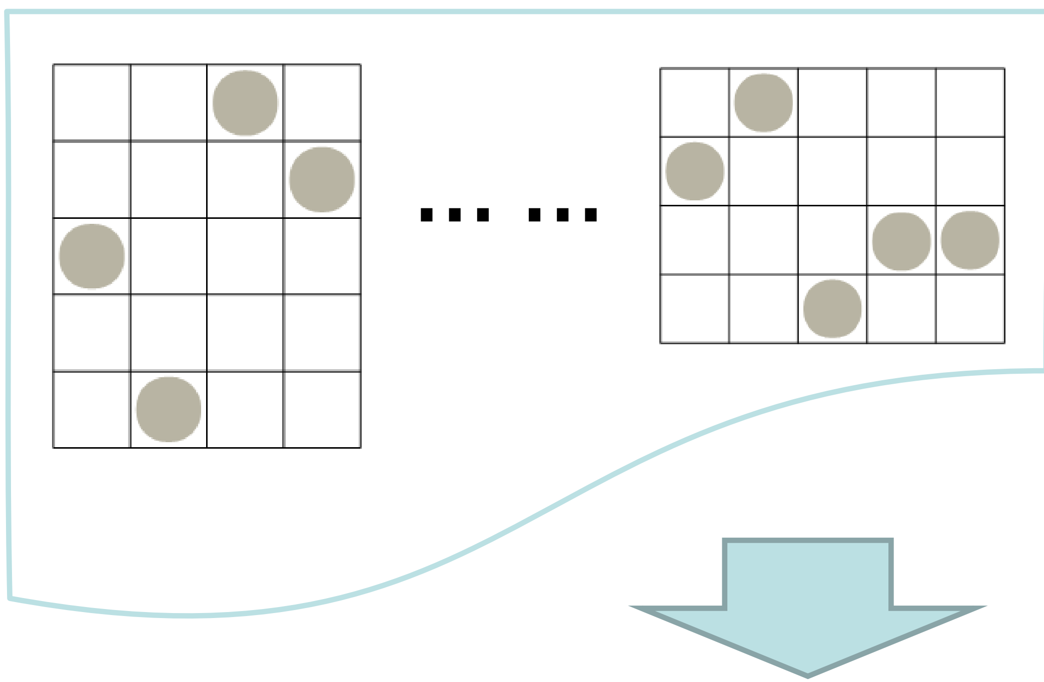
- Topic-insensitive rules are always penalized
- But common, sometime more preferable
- Sensitivity as a complement

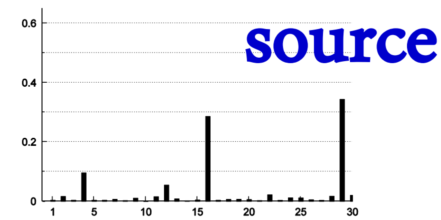
Estimation



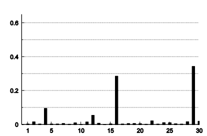


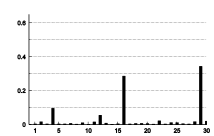
Estimation





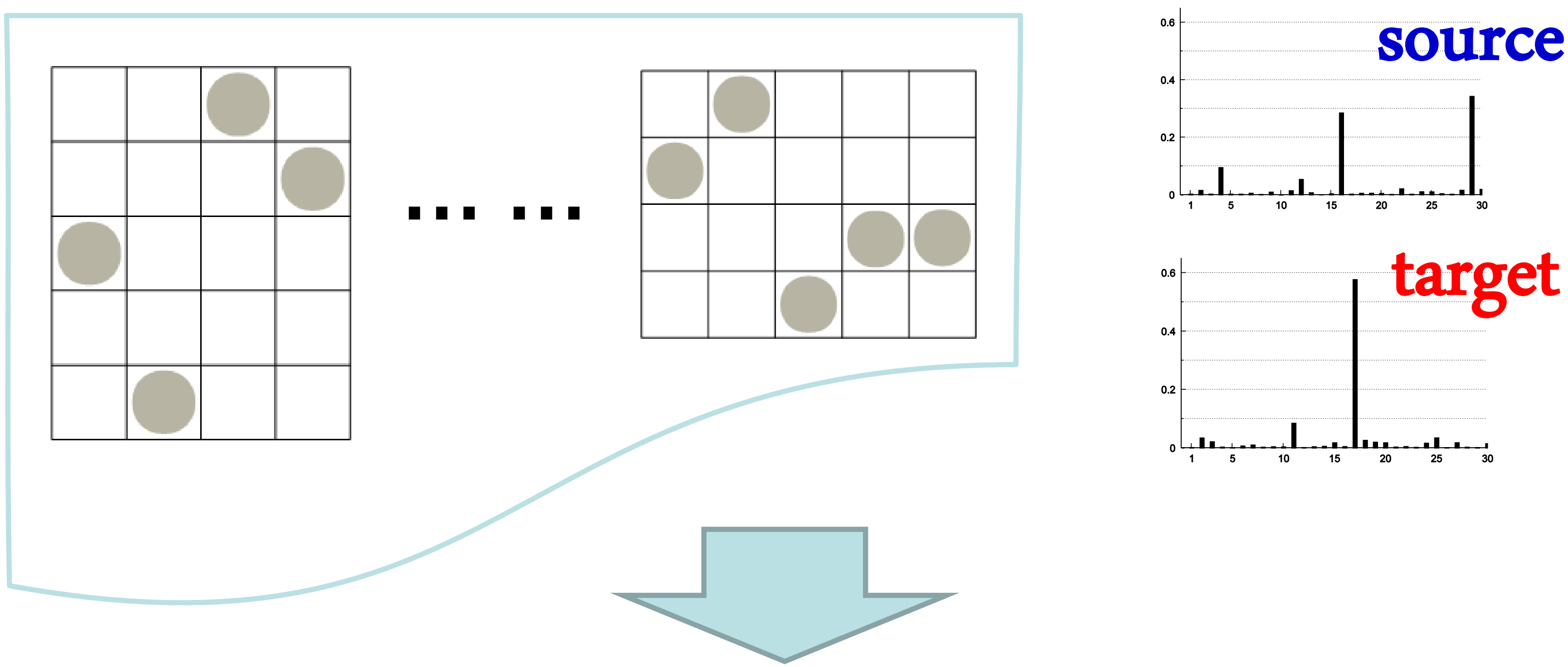
... ...

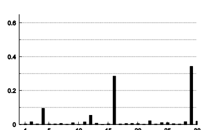
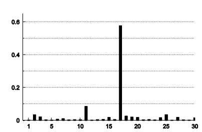
(r_1, 0.1, )

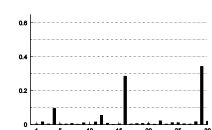
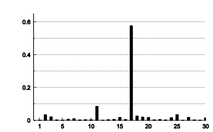
(r_2, 0.5, )

... ...

Estimation

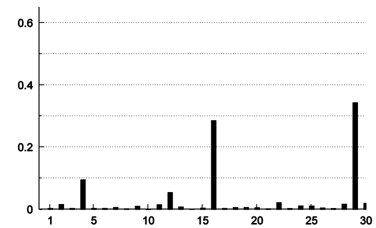


(r_1, 0.1,  , )

(r_2, 0.5,  , )

....

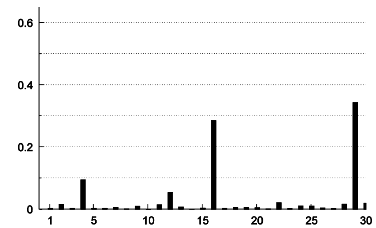
One-to-many Topic Projection



Target

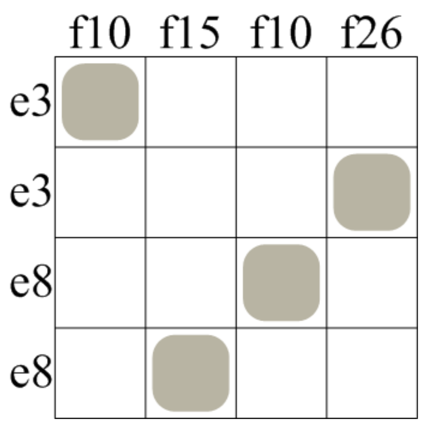
Distribution

One-to-many Topic Projection



Target Distribution

$\times$



Topic Assignment Alignment

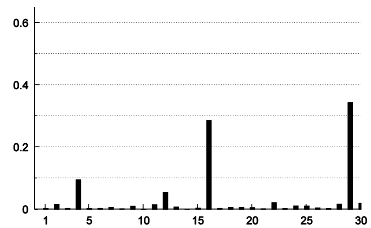
$$\begin{matrix} 0.1 & 0.4 & 0.1 & \cdots \\ 0.3 & 0.2 & 0.1 & \cdots \\ 0.4 & 0.1 & 0.3 & \cdots \\ \cdots & \cdots & \cdots & \ddots \end{matrix}$$

Topic-to-Topic Projection Matrix

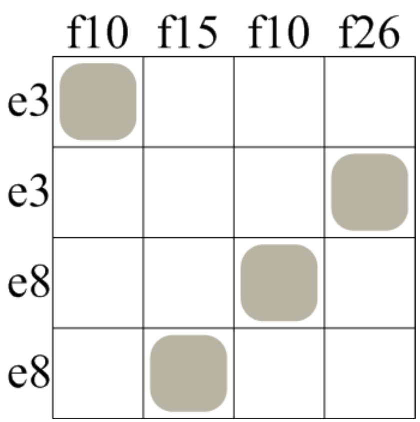
One-to-many Topic Projection

e-topic	f-topic 1	f-topic 2
enterprises	农业(agricultural)	企业(enterprise)
rural	农村(rural)	市场(market)
state	农民(peasant)	国有(state)
agricultural	改革(reform)	公司(company)
market	财政(finance)	金融(finance)
reform	社会(social)	银行(bank)
$P(z_f \mid z_e)$	0.38	0.28

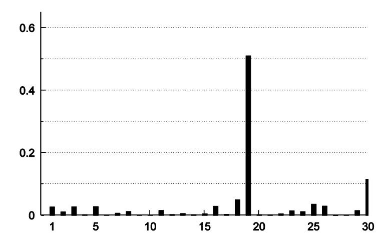
One-to-many Topic Projection



Target Distribution



$$\begin{matrix} 0.1 & 0.4 & 0.1 & \cdots \\ 0.3 & 0.2 & 0.1 & \cdots \\ 0.4 & 0.1 & 0.3 & \cdots \\ \cdots & \cdots & \cdots & \ddots \end{matrix}$$

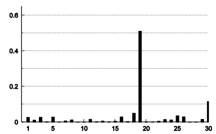
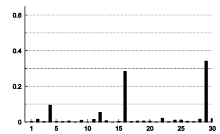
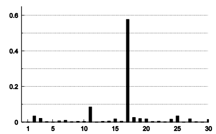
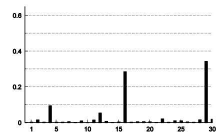
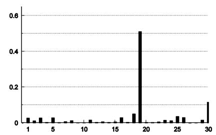
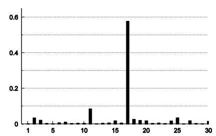


Topic Assignment Alignment

Topic-to-Topic Projection Matrix

Projected Target Distribution

Topic-based Rule Selection Model

- Similarity ( , ) source
- Similarity ( , ) target
- Sensitivity() source
- Sensitivity() target

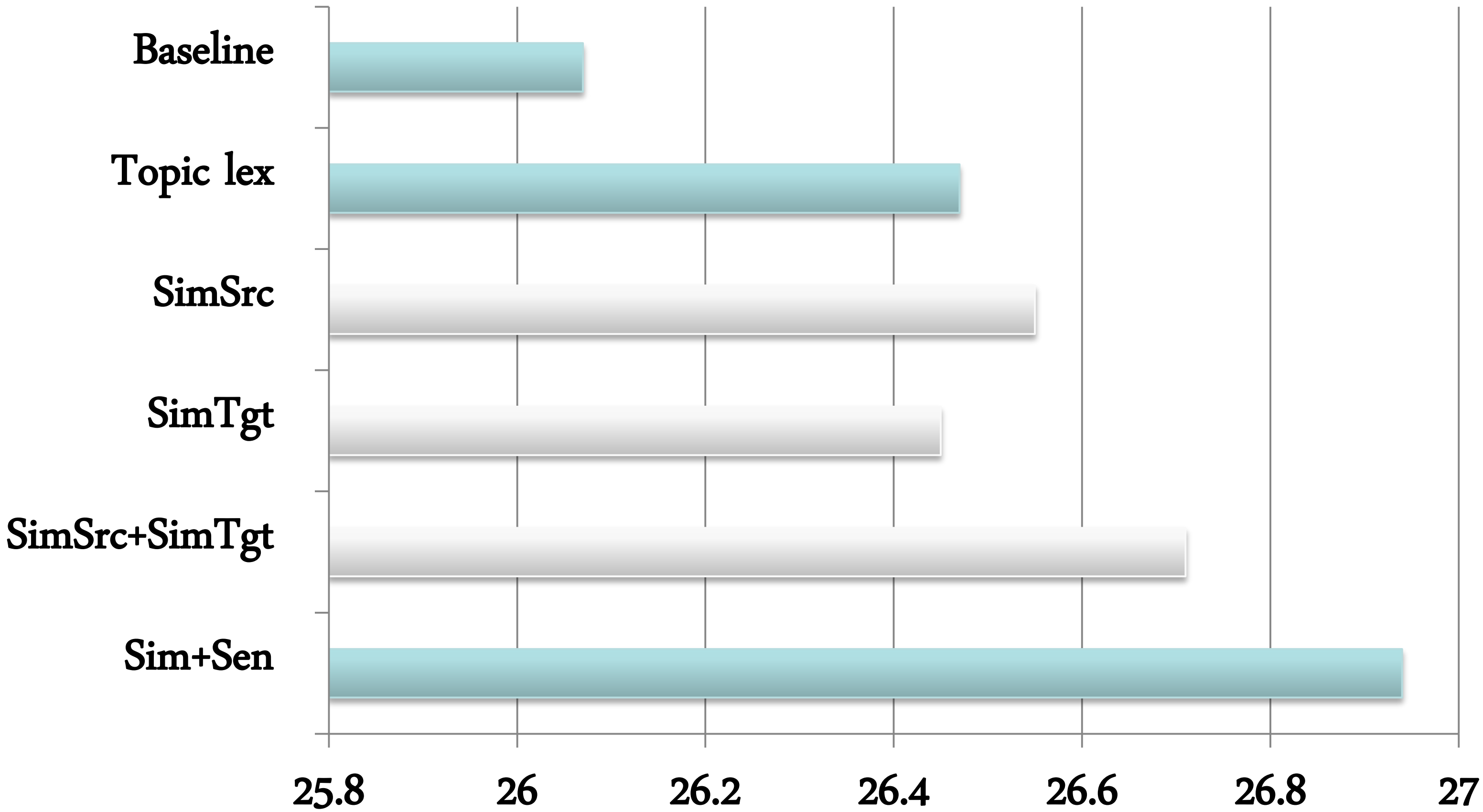
Used as four features in log-linear model for SMT

Xinyan Xiao et al. ACL 2012

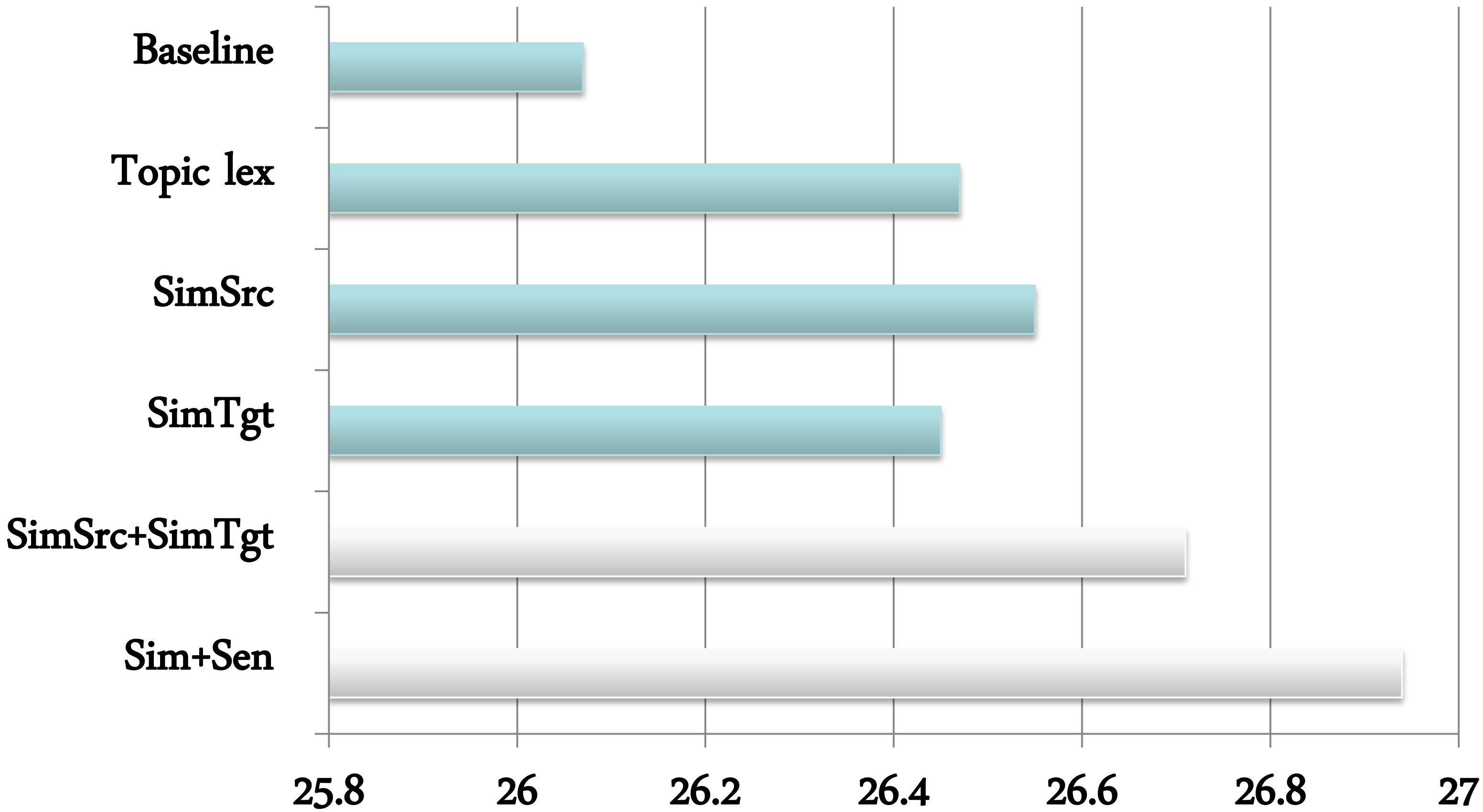
Experiment Setup

- In-house implementation of HPB model
- Topic Tool: GibbsLDA++
- Bilingual corpus: FBIS 239K sentence pairs
  - With document boundary
  - For both LDA training and rule extraction
- Report Average BLEU on test sets NIST06, NIST08

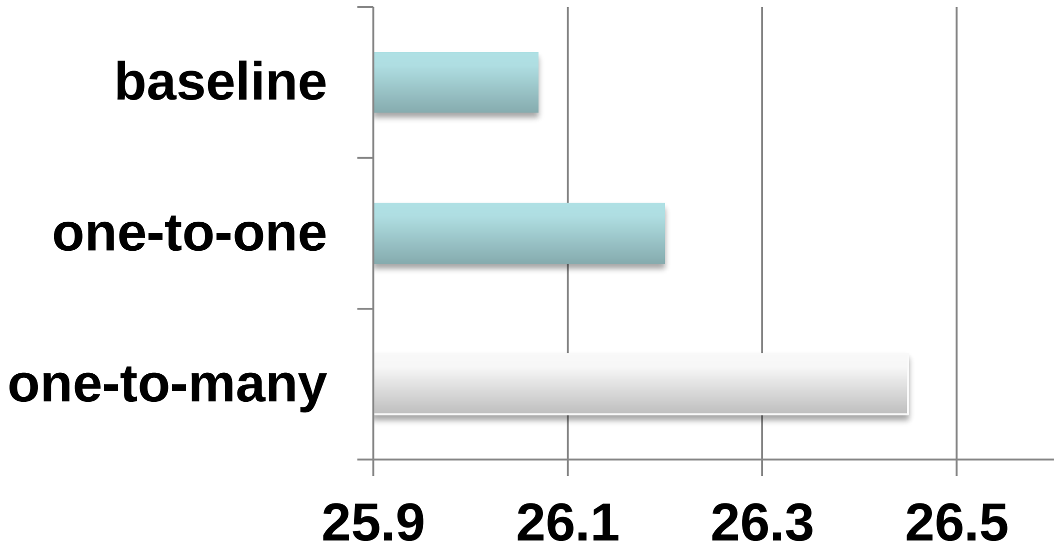
Effect of Topic Similarity Model



Effect of Sensitivity Model



One-to-many Topic Projection



Summary

- Compared with word-level WSD, our Topic-based Rule Selection Model is more effective.
- A topic similarity model and a topic sensitive model are used in both source side and target side.
- Document boundary is necessary in training corpus.

CARS Application Examples

- CARS for Bracketing Transduction Grammar
- CARS for Hierarchical Phrase-based Model
- CARS for Tree-to-String Model
- CARS using Topic Model
- CARS for Agglutinative Language Translation

Outline

- Introduction
- Context-Aware Rule-Selection
- CARS Application Examples
- Conclusion and Future Work

Conclusion

- An idea of Context-Aware Rule-Selection is proposed
- CARS is very effective on various translation models
- CARS is compatible with log-linear model for SMT
- CARS is very convenient for incorporating various context features and linguistic knowledge.

Future Work

- CARS by Semantic Role Labeling
- CARS by Coreference
- ……

Thanks!
Q & A