Adaptation for Natural Language Processing

Qun Liu, Wenbin Jiang

COLING 2014 Invited Speech

CNGL CENTRE FOR GLOBAL INTELLIGENT CONTENT
ICT 中科院计算所 INSTITUTE OF COMPUTING TECHNOLOGY,CAS

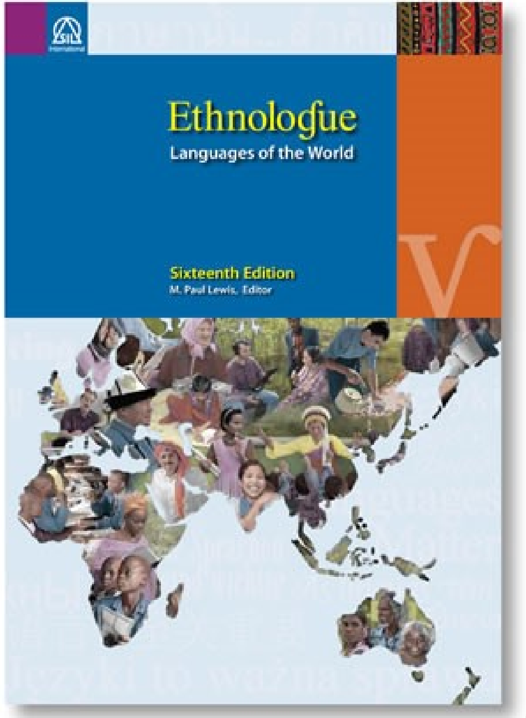
Outline

- **Introduction**
- Cross-Standard Adaptation
- Cross-Lingual Adaptation
- Experiments on Irish Processing
- Conclusion

Outline

- **Introduction**
  - **Data Scarcity Forever**
  - Existing Solutions
  - Adaptation for NLP
  - Our Contribution
- Cross-Standard Adaptation
- Cross-Lingual Adaptation
- Experiments on Irish Processing
- Conclusion

How many languages are there in the world?

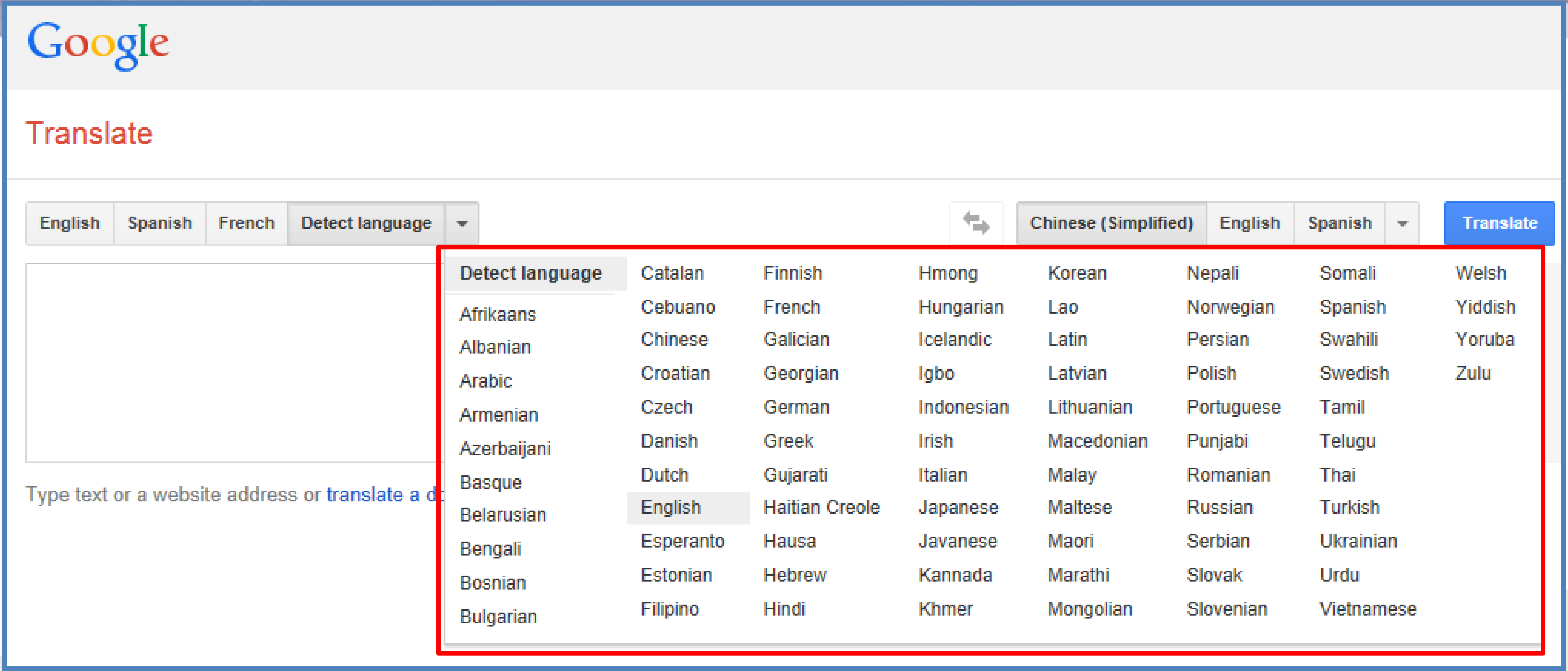


As of 2009

- At least a portion of the bible had been translated into 2,508 different languages
- The *Ethnologue* detailed classified list included 6,909 distinct languages.
- 393 languages have more than 1M speakers.

Google Translation Supports 80 Languages

CENTRE FOR GLOBAL INTELLIGENT CONTENT



NLP Tasks

- Human-annotated gold standard data is necessary for many NLP tasks:
  - Word Segmentation
  - Morphological Analysis
  - POS Tagging
  - Parsing
  - Word Sense Disambiguation (WSD)
  - Semantic Role Labelling (SRL)

Data Scarcity

To build sufficient corpora for all NLP task for all these languages is an impossible mission.

Data Scarcity will be a problem for NLP forever.

Outline

- **Introduction**
  - Data Scarcity Forever
  - **Existing Solutions**
  - Adaptation for NLP
  - Our Contribution
- Cross-Standard Adaptation
- Cross-Lingual Adaptation
- Experiments on Irish Processing
- Conclusion

Human Annotation

Advantages	Disadvantages
• High quality	• Labor intensive • Time consuming • Expensive

Crowdsourcing

Advantages	Disadvantages
• Low cost • Short development period • Public engagement	• Management • Low Consistency • Possible low quality

Unsupervised Learning

Advantages	Disadvantages
• Low cost • Good consistency	• Low performance • Does not comply with human intuition

Machine-Assisted Annotation by Active Learning

Advantages	Disadvantages
• High Quality • More Efficient	• Labor Intensive • Time Consuming • Expensive

Outline

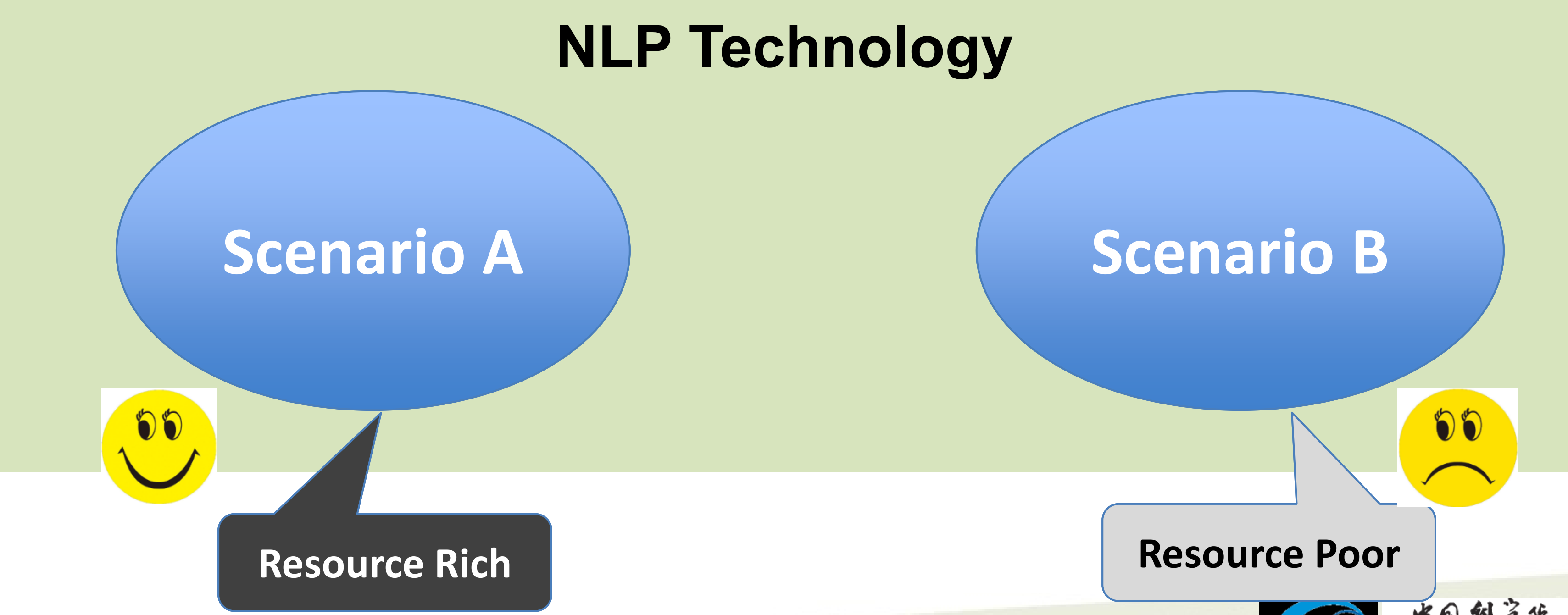
Adaptation

NLP Technology

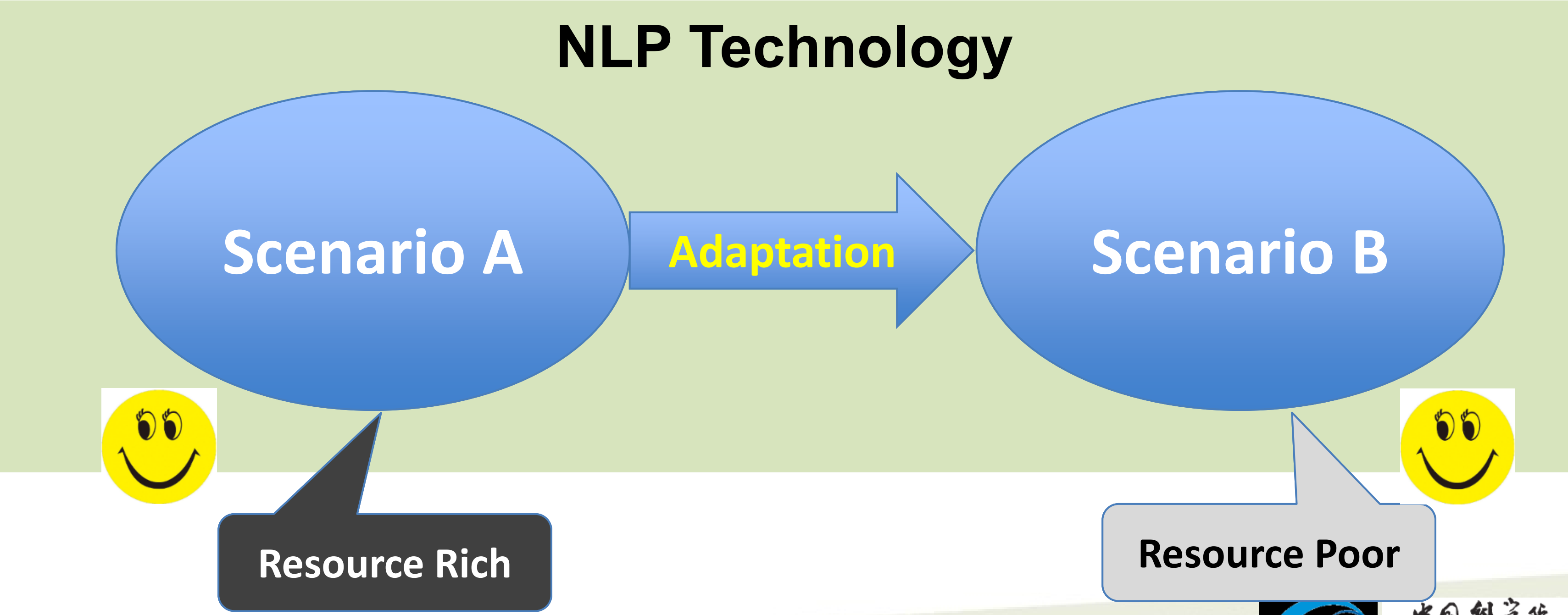
Scenario A

Scenario B

Adaptation


NLP Technology
Scenario A
Scenario B
Resource Rich
Resource Poor

Adaptation



NLP Technology
Scenario A
Adaptation
Scenario B
Resource Rich
Resource Poor

Adaptation

Adaptation is an efficient way to alleviate data scarcity problem.

Adaptation has recently attracted increasing attention.

However, it is still insufficiently researched.

Existing Adaptation Work

- **Domain Adaptation**
  - **Machine Translation**
  - **Parsing**
  - **Word Segmentation**
- **Cross-standard Adaptation**
  - **Word Segmentation**
  - **Parsing**
- **Cross-lingual Adaptation**
  - **Parsing**
  - **POS tagging**
  - **Sentiment Analysis**
- **Cross-modal Adaptation**
- **Cross-cultural Adaptation**

Intensively Researched

Developing

Emerging

Representative Work on Domain Adaptation

- Domain Adaptation for Statistical Classifiers.
  Hal Daum´e III and Daniel Marcu. In JAIR 2006
- Reranking and Self-Training for Parser Adaptation.
  David McClosky, Eugene Charniak, and Mark Johnson. In ACL 2006
- Dependency Parsing and Domain Adaptation with LR Models and Parser Ensembles.
  Kenji Sagae and Jun'ichi Tsujii. In CoNLL 2007
- Experiments in Domain Adaptation for Statistical Machine Translation.
  Philipp Koehn and Josh Schroeder. In Second Workshop on Statistical Machine Translation, 2007
- Domain Adaptation for Machine Translation by Mining Unseen Words.
  Hal Daume´ III and Jagadeesh Jagarlamudi. In ACL 2011

Representative Work on Cross-standard Adaptation

- Automatic annotation of the penn treebank with lfg f-structure information.
  Aoife Cahill, Mairead McCarthy, Josef van Genabith and Andy Way. In Proceedings of the LREC Workshop, 2002
- Adaptive chinese word segmentation.
  Jianfeng Gao, Andi Wu, Mu Li, Chang-Ning Huang, Hongqiao Li, Xinsong Xia, and Haowei Qin. In Proceedings of ACL, 2004
- CCGbank: a corpus of CCG derivations and dependency structures extracted from the penn treebank.
  Julia Hockenmaier and Mark Steedman. In Computational Linguistics, 2007

Representative Work on Cross-lingual Adaptation

- Bootstrapping parsers via syntactic projection across parallel texts. *Rebecca Hwa, Philip Resnik, Amy Weinberg, Clara Cabezas, and Okan Kolak. In Natural Language Engineering, 2005*
- Parser adaptation and projection with quasi-synchronous grammar features. *David Smith and Jason Eisner. In Proceedings of EMNLP, 2009*
- Unsupervised part-of-speech tagging with bilingual graph-based projections. *Dipanjan Das and Slav Petrov. In Proceedings of ACL, 2011*
- Dependency grammar induction via bitext projection constraints. *Ganchev, Kuzman, Jennifer Gillenwater, and Ben Taskar. In Proceedings of ACL, 2009*

COLING 2014 Adaptation Papers

1. **Cross-lingual Coreference Resolution of Pronouns**
   Michal Novak and Zdenek Zabokrtsky
2. **Cross-lingual Discourse Relation Analysis: A corpus study and a semi-supervised classification system**
   Junyi Jessy Li, Marine Carpuat and Ani Nenkova
3. **Cross-Topic Authorship Attribution: Will Out-Of-Topic Data Help?**
   Upendra Sapkota, Thamar Solorio, Manuel Montes, Steven Bethard and Paolo Rosso
4. **Rediscovering Annotation Projection for Cross-Lingual Parser Induction**
   Jörg Tiedemann
5. **Soft Cross-lingual Syntax Projection for Dependency Parsing**
   Zhenghua Li, Min Zhang and Wenliang Chen
6. **Dynamically Integrating Cross-Domain Translation Memory into Phrase-Based Machine Translation during Decoding**
   Kun Wang, Chengqing Zong and Keh-Yih Su
7. **Enriching Wikipedia's Intra-language Links by their Cross-language Transfer**
   Takashi Tsunakawa, Makoto Araya and Hiroyuki Kaji
8. **Global methods for crosslingual semantic role and predicate labelling**
   Lonneke van der Plas, Marianna Apidianaki and chenhua chen
9. **Predicting Machine Translation Quality Estimation Across Domains**
   José G. C. de Souza, Marco Turchi and Matteo Negri

Outline

- **Introduction**
  - Data Scarcity Forever
  - Existing Solutions
  - Adaptation for NLP
  - **Our Contribution**
- Cross-Standard Adaptation
- Cross-Lingual Adaptation
- Experiments on Irish Processing
- Conclusion

Problem

Cross-Standard

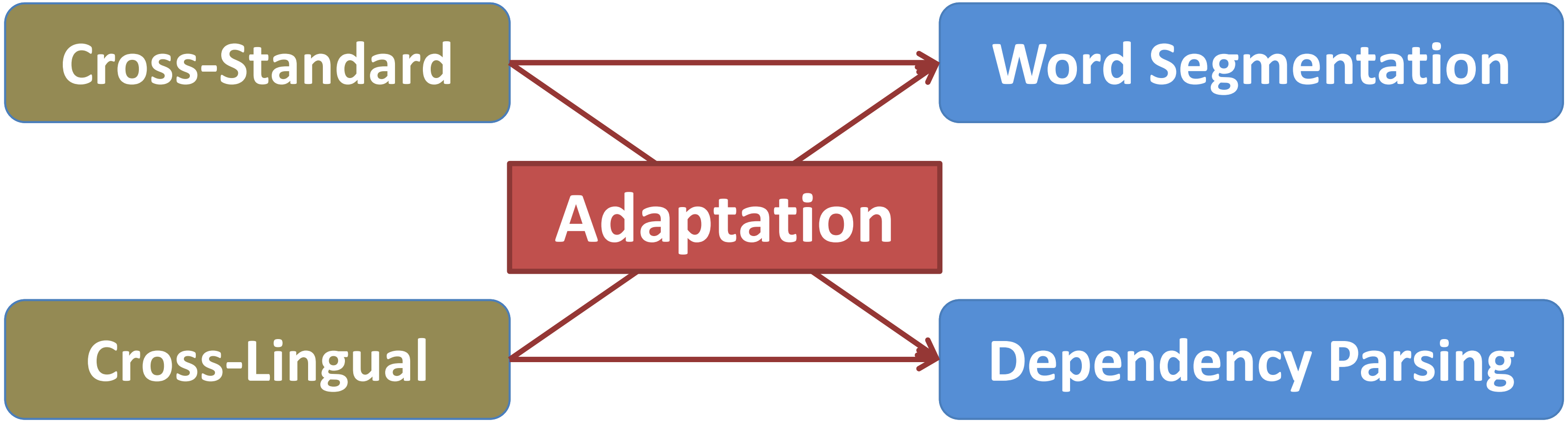
Cross-Lingual

Our Contribution

Conditional Mapping for Cross-standard Adaptation

Decomposed Projection for Cross-lingual Adaptation

Problem


Cross-Standard
Word Segmentation
Adaptation
Cross-Lingual
Dependency Parsing

Chinese Word Segmentation

- Input:

➢ 今天是星期三。

- Output:

➢ 今天 / 是 / 星期三 / 。

Chinese Word Seg. by Character Annotation

- Instead of directly inserting delimiters between words, we annotate each character with a label indicating the position of the character in a word:
  - 今/B 天/E 是/S 星/B 期/M 三/E 。/S
    - B: The first character in a word
    - M: The middle character in a word
    - E: The last character in a word
    - S: The single character is a word

Chinese Word Seg. by Character Annotation

1. Calculate the probability of all the characters to be annotated as each of the labels:

$$p(t_i\,|C_i,\ s{=}C_1C_2\ldots C_n),\ i=1,\ldots,\mathrm{n},\ t_i\in\{\mathrm{B,M,E,S}\}$$

2. A Viterbi algorithm is used to find the best legal path and the segmentation is generated.

$$\mathrm{argmax}(t_1\ldots t_n)\ \mathrm{product}(i)\ p(t_i\,|C_i,\ s{=}C_1C_2\ldots C_n)$$

中国科学院计算技术研究所

Chinese Word Seg. by Character Annotation

- So the segmentation problem is converted to a character classification problem.
- Classification algorithms: ME, Perceptron, CRF， …
- Features: current character: C_0, predicted label: T_0
  - $C_n T_0 (n = -2, -1, 0, 1, 2)$： current character
  - $C_n C_{n+1} T_0 (n = -2, -1, 0, 1)$： character bi-gram
  - $C_{-1} C_1 T_0$： neighbor characters
  - $D(C_0) T_0$： if the current character is a digit
  - $A(C_0) T_0$： if the current character is a Latin letter
  - $P(C_0) T_0$： if the current character is a punctuation

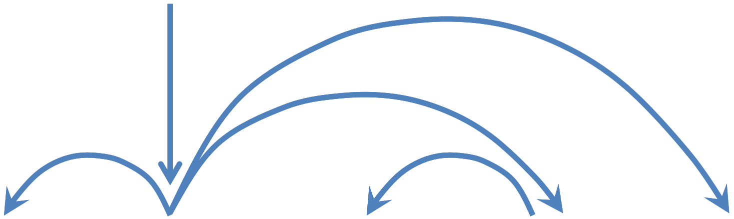
中国科学院计算技术研究所

Dependency Parsing

- Input:

➢ 我 是 一个 学生 。

- Output：



➢ 我 是 一个 学生 。

Dep. Parsing by Maximum Spanning Tree

1. Calculate the probability of if there is a dependency relation between all the word pairs:

$p(w_i \rightarrow w_j \mid s=w_1w_2\ldots w_n)$, $i, j = 1,\ldots,n$

p-1 p p+1 …r …c-1…c c+1

2. A Viterbi algorithm is used to find the best legal path and the segmentation is generated.

argmax(any spanning tree T)

pruduct($(i,j)\in T$) $p(w_i \rightarrow w_j \mid s=w_1w_2\ldots w_n)$)

Dep. Parsing by Maximum Spanning Tree

- Thus the dependency parsing problem is converted to a word pair classification problem
- Classification algorithms: ME, Perceptron , …
- Features:

Pword, Ppos
Pword
Ppos
Cword, Cpos
Cword
Cpos

Pword, Ppos, Cword, Cpos
Ppos, Cword, Cpos
Pword, Cword, Cpos
Pword, Ppos, Cpos
Pword, Ppos, Cword
Pword, Cword
Ppos, Cpos

Pword, Bpos, Cpos

Ppos,Ppos+1,Cpos-1, Cpos
Ppos-1, Ppos, Cpos-1, Cpos
Ppos, Ppos+1, Cpos, Cpos+1

p-1 p p+1 …r …c-1…c c+1

Outline

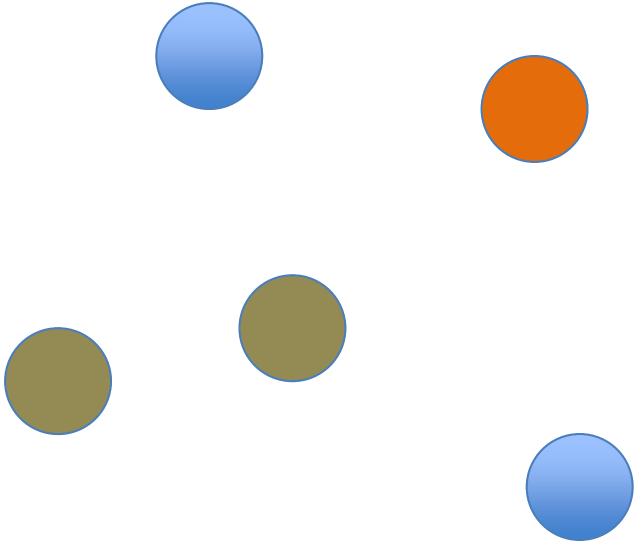
Outline

- Introduction
- **Cross-Standard Adaptation**
  - **Conditional Mapping**
  - Word Segmentation
  - Dependency Parsing
- Cross-Lingual Adaptation
- Experiments on Irish Processing
- Conclusion

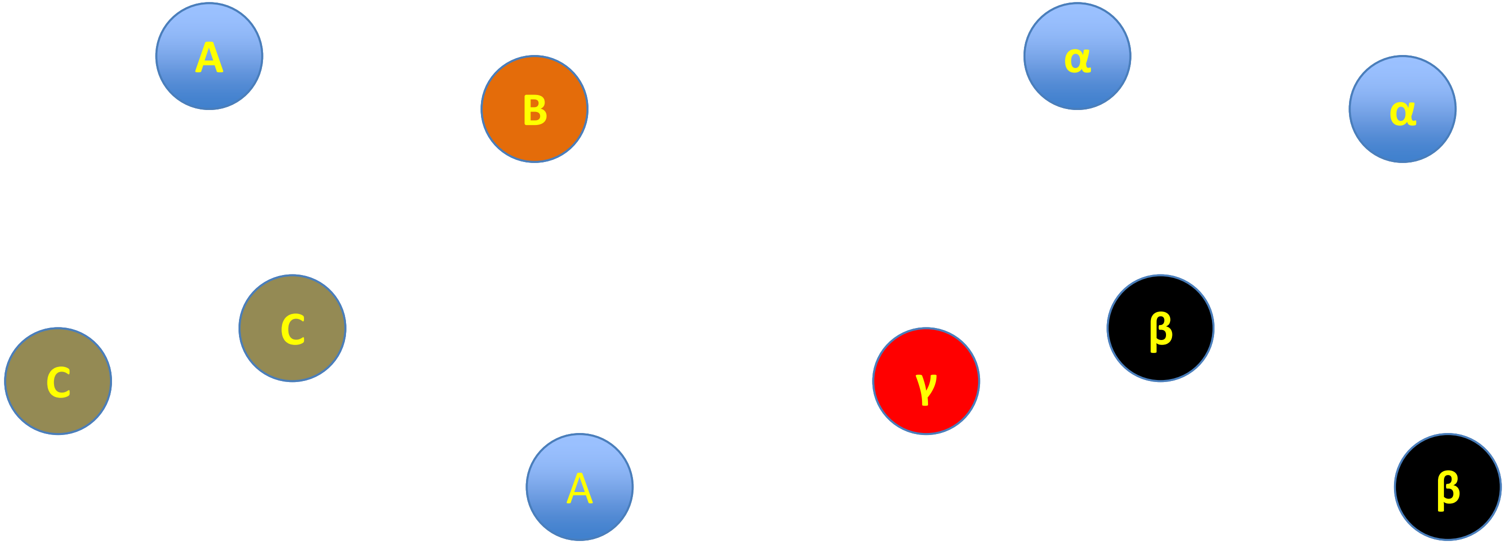
Conditional Mapping

for Cross-standard Adaptation

Cross-standard Adaptation

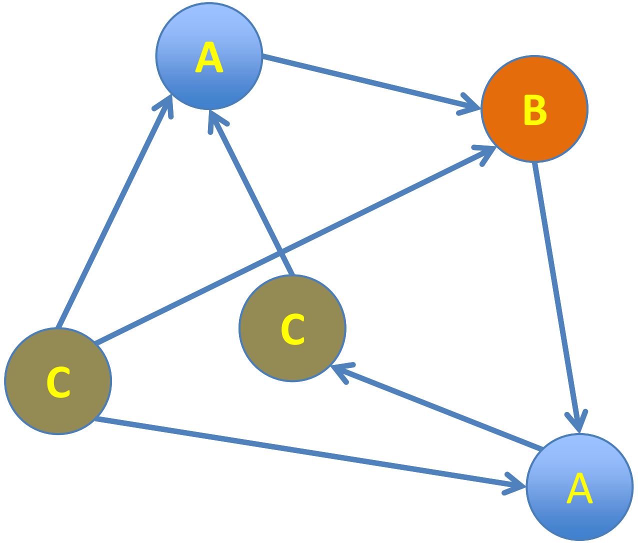


Cross-standard Adaptation

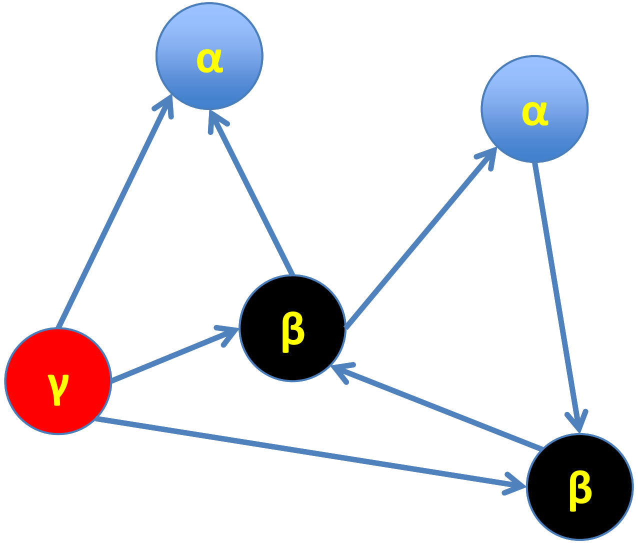


Annotation Standard 1 **Annotation Standard 2**

Cross-standard Adaptation



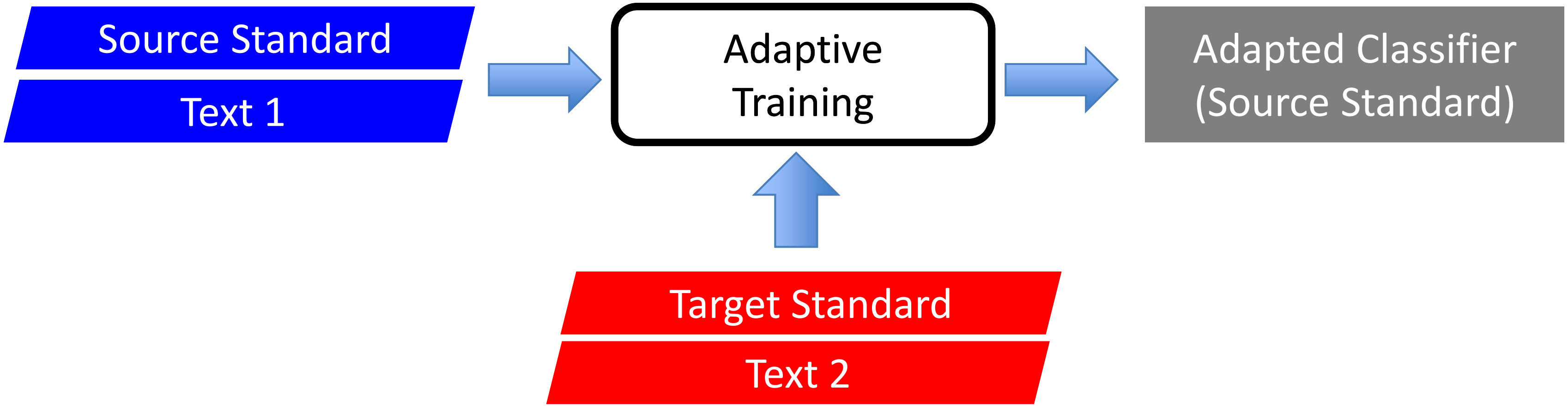
Annotation Standard 1



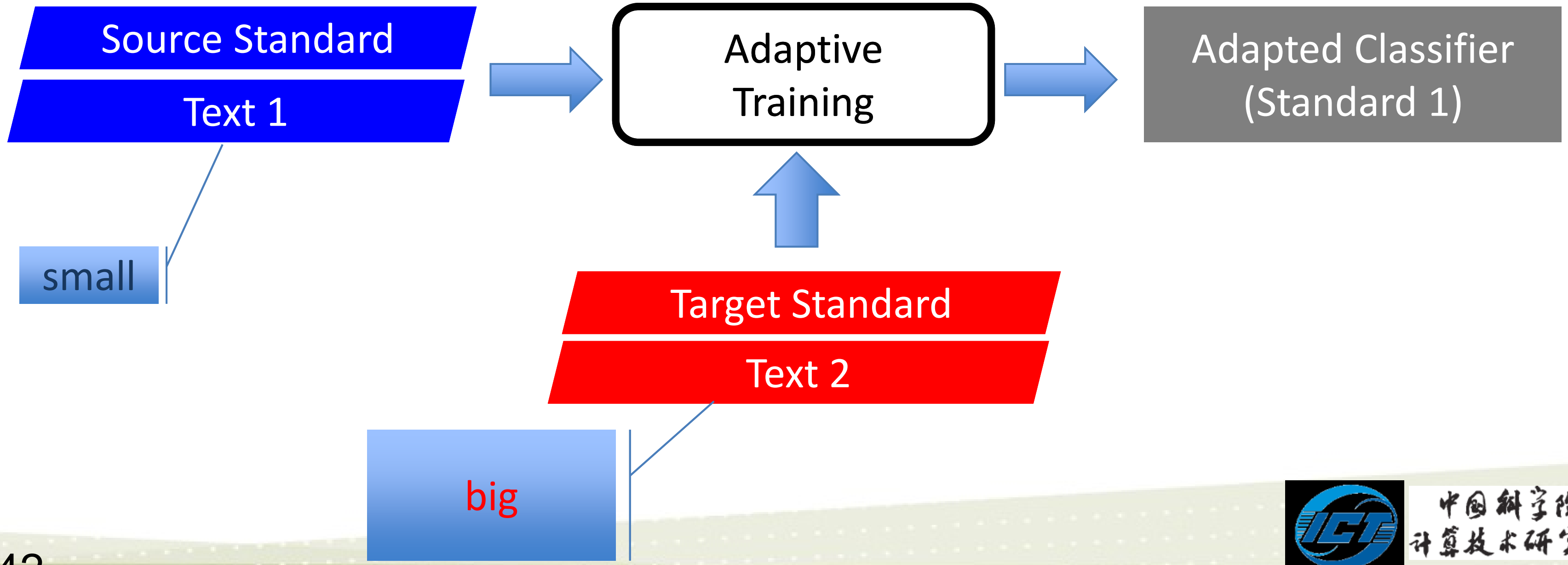
Annotation Standard 2

Cross-standard Adaptation

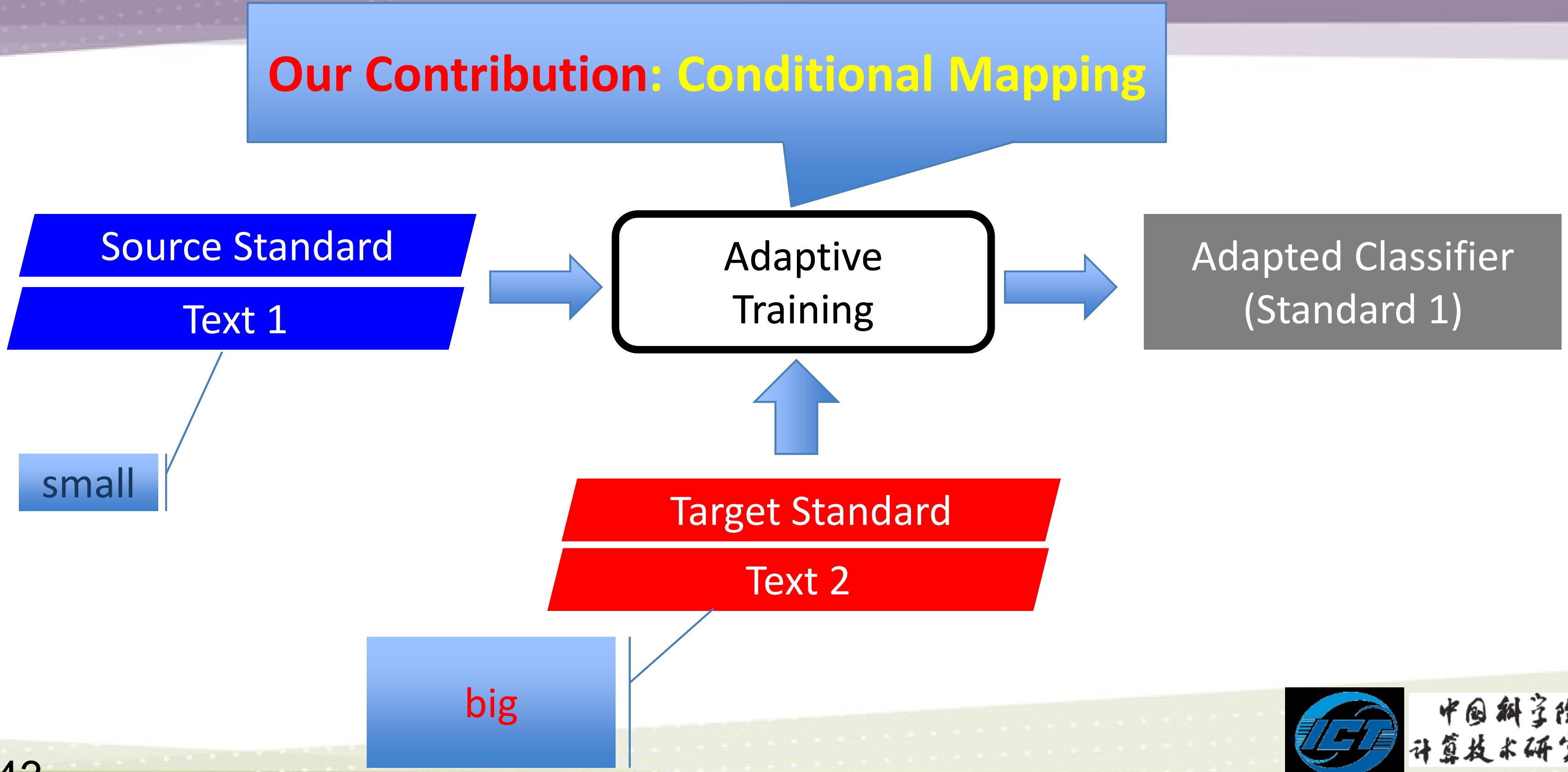
Cross-standard Adaptation



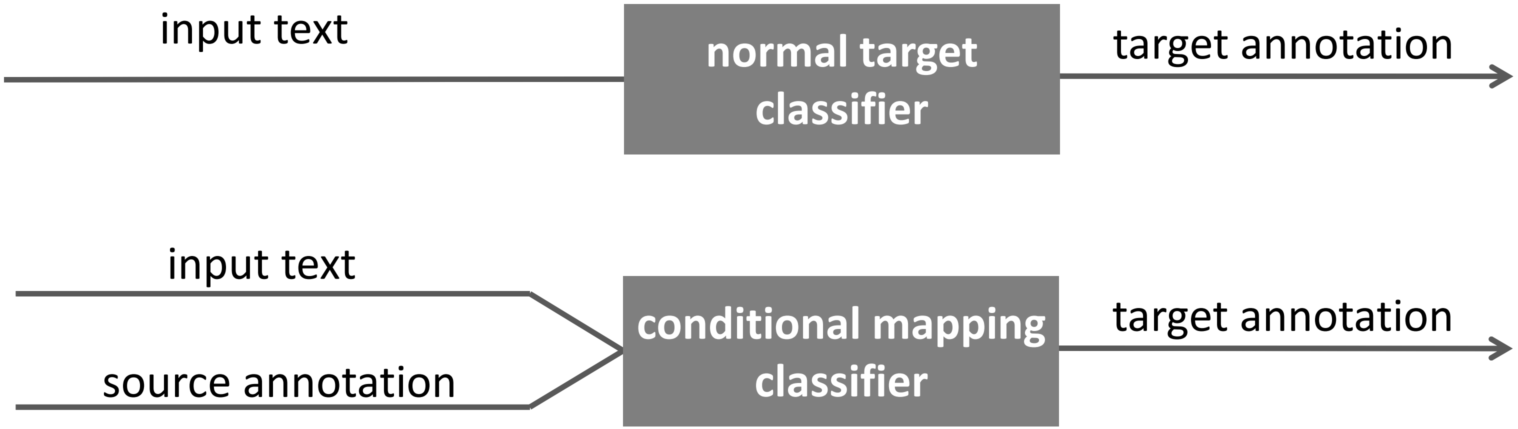
Cross-standard Adaptation



Cross-standard Adaptation

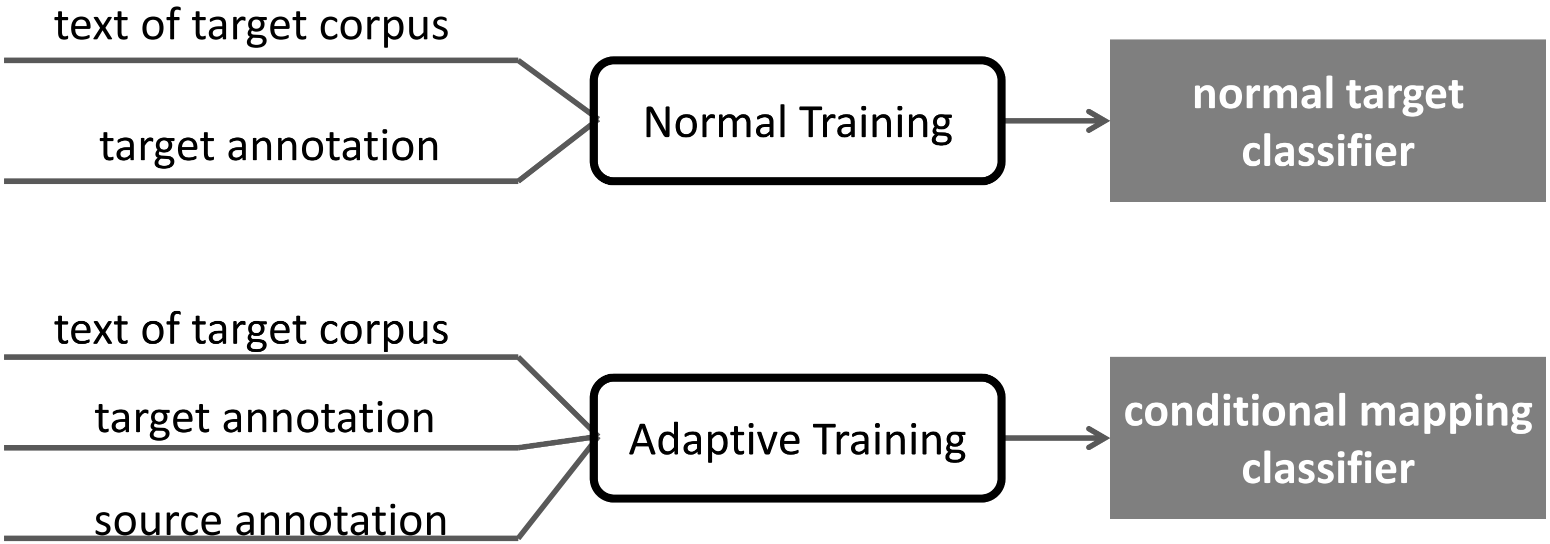


Conditional Mapping Classifier



P(target annotation | context, source annotation)

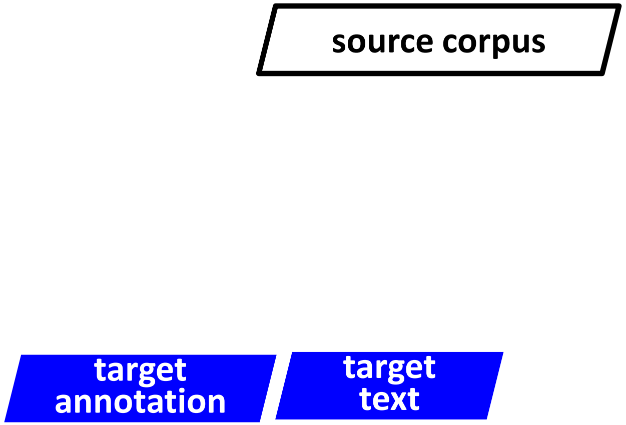
Conditional Mapping Classifier



Conditional Mapping Classifier Training

- Unfortunately, a parallel annotated corpus with gold annotations does not exist

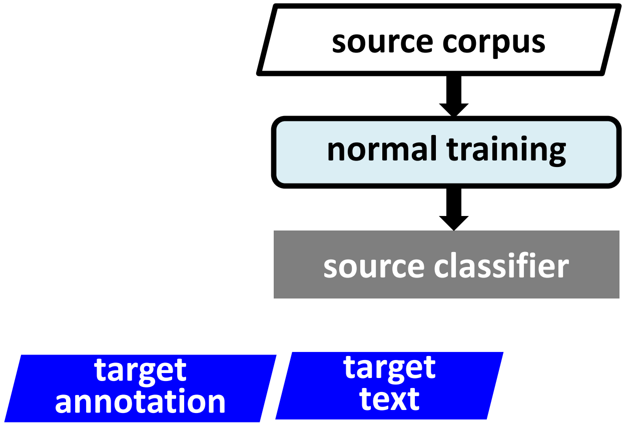
- Build a noisy one automatically



Conditional Mapping Classifier Training

- Unfortunately, a parallel annotated corpus with gold annotations will not exist

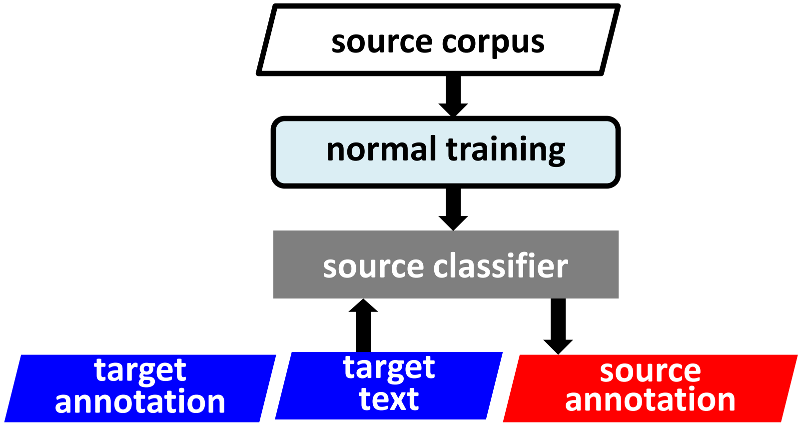
- Build a noisy one automatically



Conditional Mapping Classifier Training

- Unfortunately, a parallel annotated corpus with gold annotations will not exist

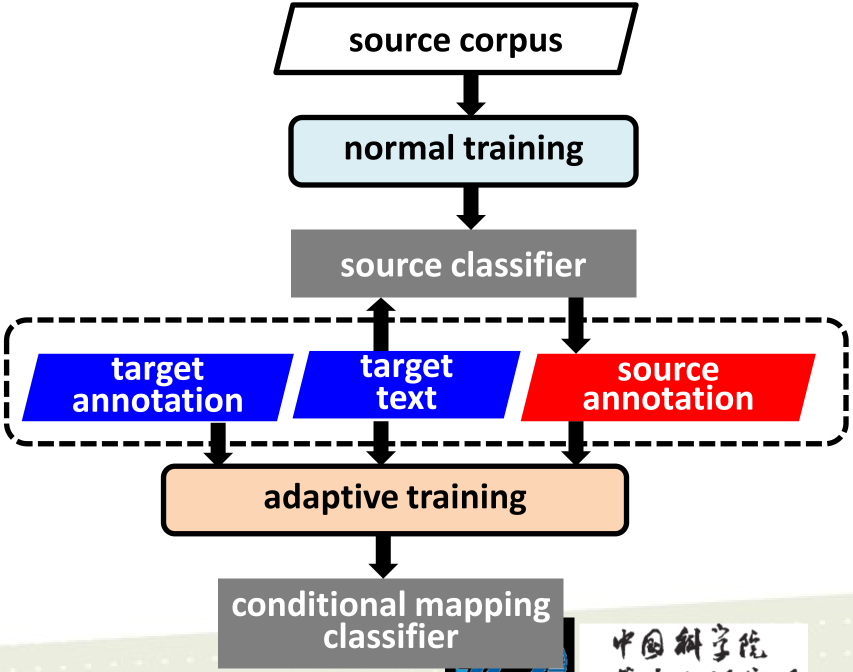
- Build a noisy one automatically



Conditional Mapping Classifier Training

- Unfortunately, a parallel annotated corpus with gold annotations will not exist

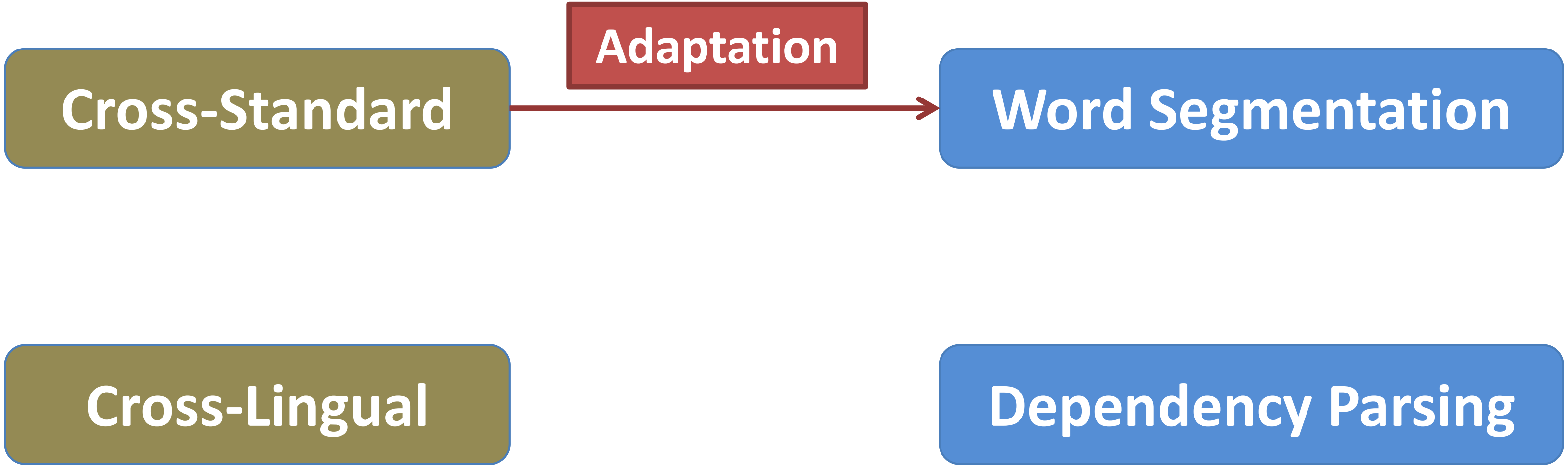
- Build a noisy one automatically



Outline

- Introduction
- **Cross-Standard Adaptation**
  - Conditional Mapping
  - **Word Segmentation**
  - Dependency Parsing
- Cross-Lingual Adaptation
- Experiments on Irish Processing
- Conclusion

Problem


Cross-Standard
Adaptation
Word Segmentation
Cross-Lingual
Dependency Parsing

Cross-standard Adaptation for Word Segmentation

- There are several annotation schemes for Chinese word segmentation, corresponding to different corpora

7M words

People’s Daily Corpus
Peking University

5M words

Sinica Corpus
Academia Sinica

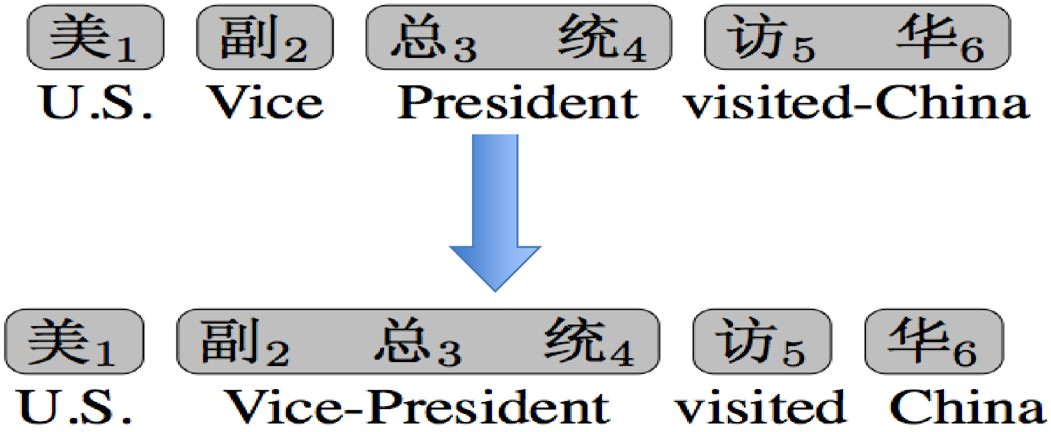
1M words

Penn Chinese Treebank
University of Pennsylvania

Cross-standard Adaptation for Word Segmentation

- Cross-standard adaptation for word segmentation aims to transform a word segmentation corpus from one annotation style to another

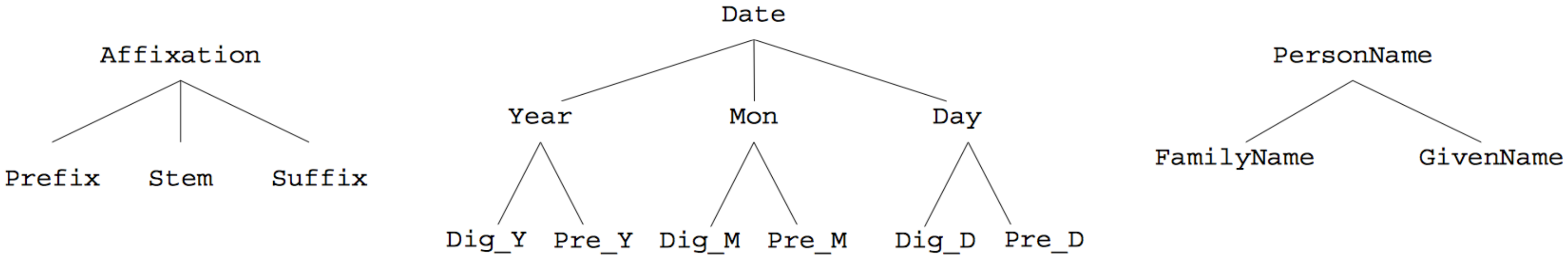
People's Daily



Penn

Previous Work

- Hand-crafted templates with error-driven learning (Gao et al., 2004)



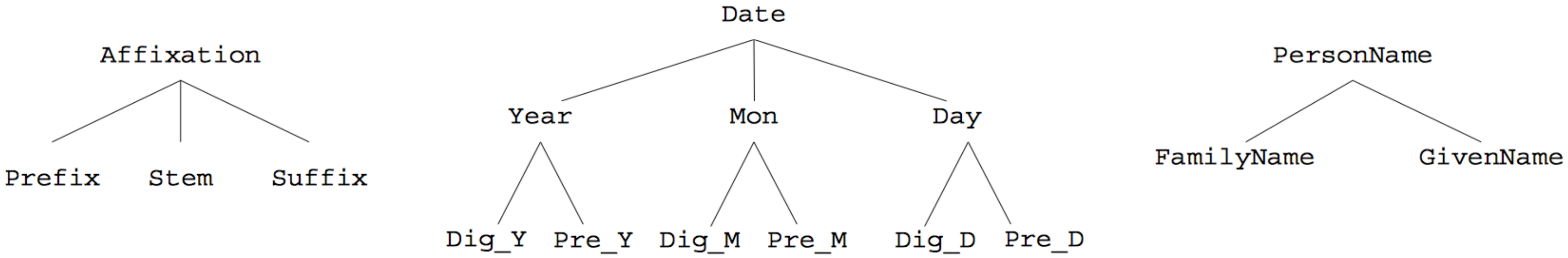
Condition: 'Affixation'
Actions: Insert a boundary between 'Prefix' and 'Stem'…

Condition: 'Date'
Actions: Insert a boundary between 'Year' and 'Mon' …

Condition: 'PersonName'
Actions: Insert a boundary between 'FamilyName' and 'GivenName'…

Previous Work

- Hand-crafted templates with error-driven learning (Gao et al., 2004)



Condition: 'Affixation'
Actions: Insert a boundary between 'Prefix' and 'Stem'…

Condition: 'Date'
Actions: Insert a boundary between 'Year' and 'Mon' …

Condition: 'PersonName'
Actions: Insert a boundary between 'FamilyName' and 'GivenName'…

Our Solution – Traditional Classifier

target anno.
text

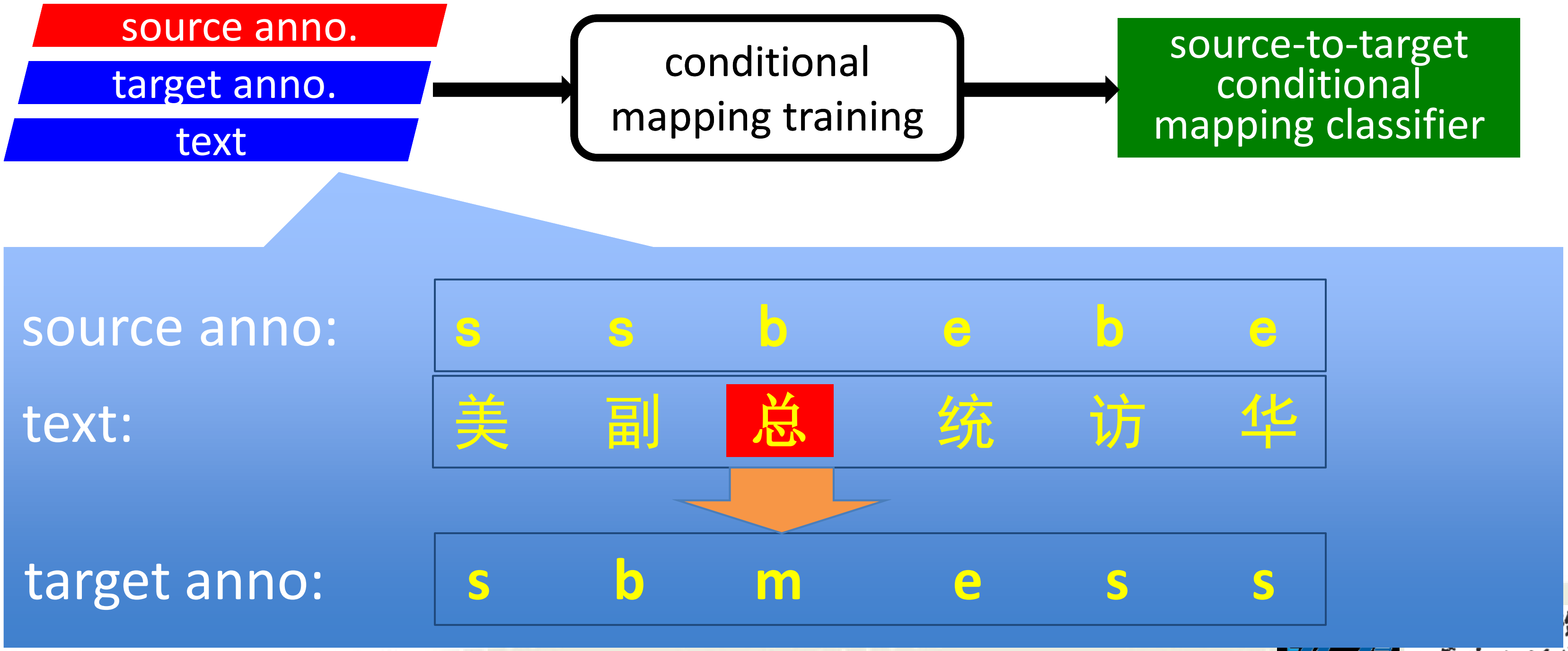
traditional classifier training

Traditional classifier

text: 美 副 总 统 访 华

target anno: s b m e s s

Our Solution – Conditional Mapping



Features

Type	Templates	Instances
n-gram	C_{-2}	C_{-2}=美
	C_{-1}	C_{-1}=副
	C_0	C_0=总
	C_1	C_1=统
	C_2	C_2=访
	$C_{-2}C_{-1}$	$C_{-2}C_{-1}$=美副
	$C_{-1}C_0$	$C_{-1}C_0$=副总
	C_0C_1	C_0C_1=总统
	C_1C_2	C_1C_2=统访
	$C_{-1}C_1$	$C_{-1}C_1$=副统
function	$Pu(C_0)$	$Pu(C_0)$=true
	$T(C_{-2:2})$	$T(C_{-2:2})$=4444

Features follow (Ng & Low 2004]

Features

Type	Templates	Instances
n-gram	C_{-2}	C_{-2}=美
	C_{-1}	C_{-1}=副
	C_0	C_0=总
	C_1	C_1=统
	C_2	C_2=访
	$C_{-2}C_{-1}$	$C_{-2}C_{-1}$=美副
	$C_{-1}C_0$	$C_{-1}C_0$=副总
	C_0C_1	C_0C_1=总统
	C_1C_2	C_1C_2=统访
	$C_{-1}C_1$	$C_{-1}C_1$=副统
function	$Pu(C_0)$	$Pu(C_0)$=true
	$T(C_{-2:2})$	$T(C_{-2:2})$=4444

Type	Templates	Instances
n-gram	C_{-2}	C_{-2}=美
	C_{-1}	C_{-1}=副
	C_0	C_0=总
	C_1	C_1=统
	C_2	C_2=访
	$C_{-2}C_{-1}$	$C_{-2}C_{-1}$=美副
	$C_{-1}C_0$	$C_{-1}C_0$=副总
	C_0C_1	C_0C_1=总统
	C_1C_2	C_1C_2=统访
	$C_{-1}C_1$	$C_{-1}C_1$=副统
Function	$Pu(C_0)$	$Pu(C_0)$=true
	$T(C_{-2:2})$	$T(C_{-2:2})$=4444

Experiment Setup

- Target corpus:
  Penn Chinese Treebank 5.0

1M words

美$_1$ | 副$_2$ 总$_3$ 统$_4$ | 访$_5$ | 华$_6$
U.S. | Vice-President | visited | China

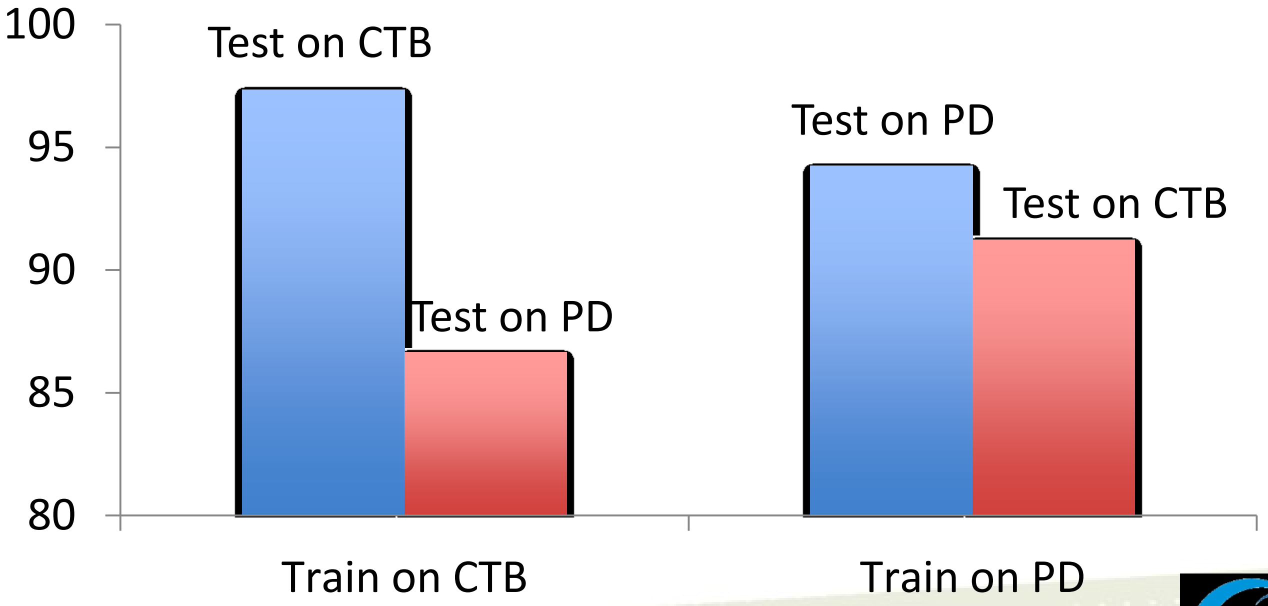
- Source corpus:
  People's Daily

7M words

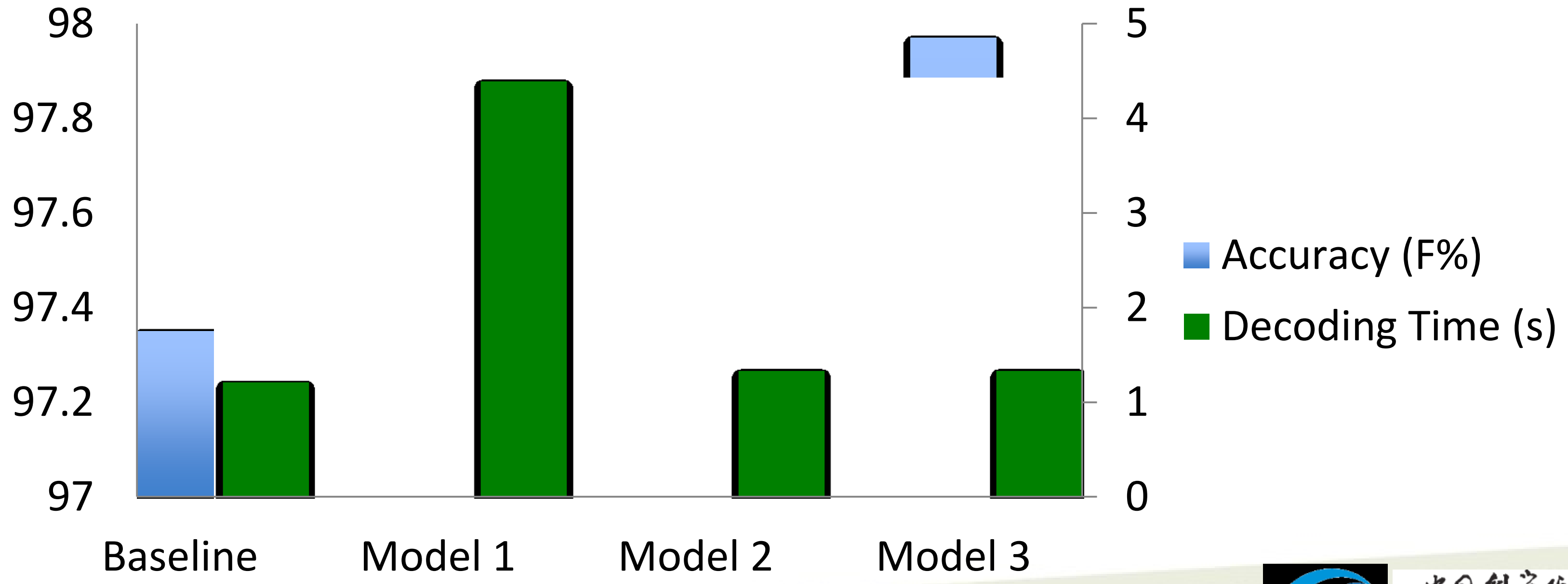
美$_1$ | 副$_2$ | 总$_3$ 统$_4$ | 访$_5$ 华$_6$
U.S. | Vice | President | visited-China

- Classifier:
  Averaged perceptron

Baseline Models



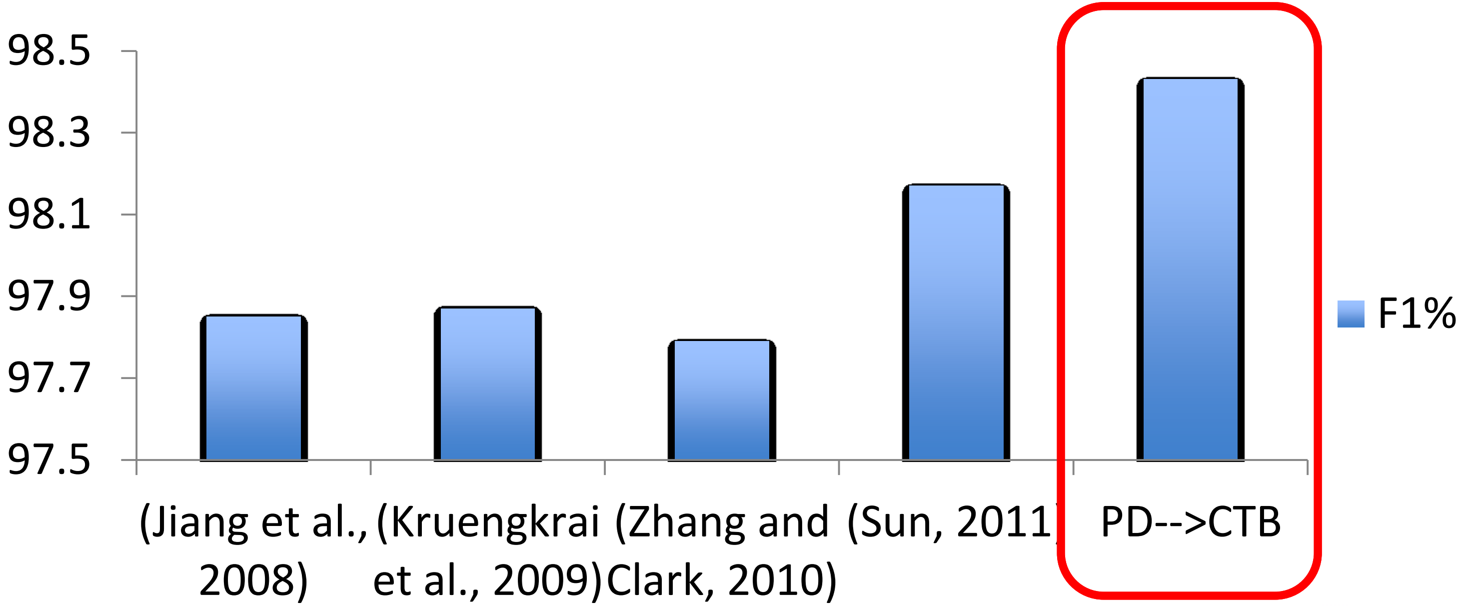
Annotation Adaptation for Word Segmentation



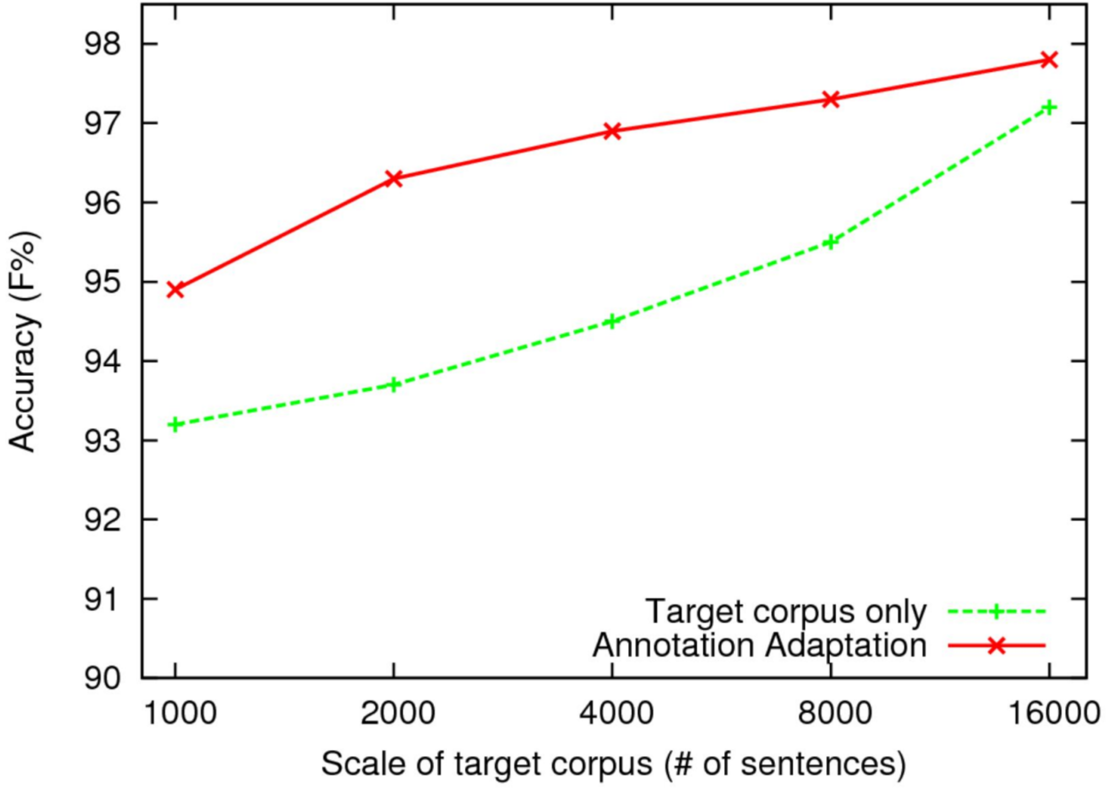
Our Work vs. Non-adaptation Work

Representative Previous Work	Model	Features	Adaptation
(Jiang et al., 2008)	Cascaded	Local + Non-local	No
(Zhang and Clark, 2010)	Single	Local + Non-local	No
(Sun, 2011)	Cascaded	Local + Non-local	No
Our Work	Single	Local	Yes

Our Work vs. Non-adaptation Work



Performance wrt #sentence



Our Work vs. Previous Adaptation Work

	Method	Automatic/Manual
(Gao et al., 2004)	Rule-based + statistical	Semi-automatic
Our Work	Statistical	Automatic

Publications

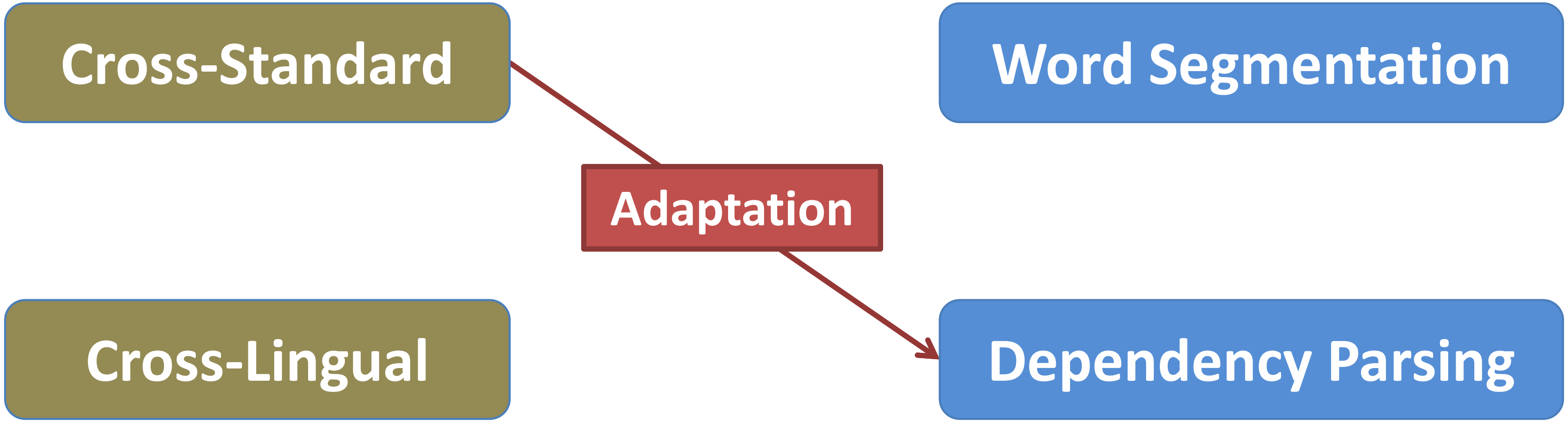
- Wenbin Jiang, Liang Huang, and Qun Liu. 2009. Automatic Adaptation of Annotation Standards: Chinese Word Segmentation and POS Tagging -- A Case Study. In *Proceedings of ACL-IJCNLP 2009*, Singapore, August.

- Wenbin Jiang, Yajuan Lü, Liang Huang and Qun Liu. 2014. Automatic Adaptation of Annotations. To appear in *Computational Linguistics*.

Outline

- Introduction
- **Cross-Standard Adaptation**
  - Conditional Mapping
  - Cross-standard Adaptation
  - **Dependency Parsing**
- Cross-Lingual Adaptation
- Experiments on Irish Processing
- Conclusion

Problem


Cross-Standard
Word Segmentation
Adaptation
Cross-Lingual
Dependency Parsing

Cross-standard Adaptation for Dependency Parsing

- There are also several popular grammatical theories for Chinese dependency parsing

1M words

Chinese Penn Treebank
University of Pennsylvania

1M words

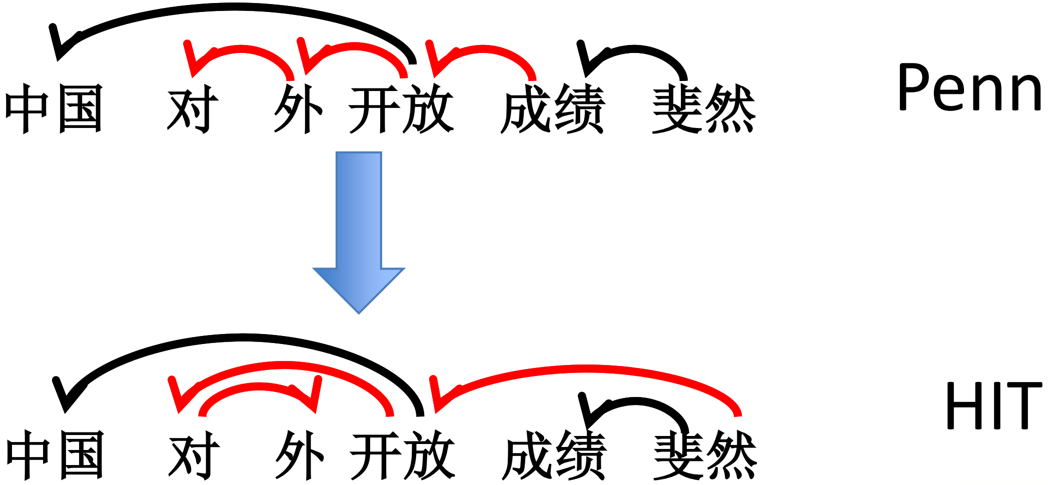
Tsinghua Treebank
Tsinghua University

0.3M words

Semantic Dependency Treebank
Harbin Institute of Technology

Cross-standard Adaptation for Dependency Parsing

- Cross-standard adaptation for dependency parsing aims to transform a treebank from one annotation style to another



Previous Work

CENTRE FOR GLOBAL INTELLIGENT CONTENT

- Hand-crafted rules for tree transformation (Cahill et al., 2002; Hockenmaier and Steedman, 2007)

designing rules for tree transformation → tree transformation

```
     s:A
    /   \
np:B     vp:C
 |        |          =>    dom(A,B), dom(A,C),
John      v:D              dom(C,D), ..
          |                pre(B,C),
         left              cat(A,s), cat(C,vp),
                           cat(D,v), ..
```

```
dom(X,Y), dom(X,Z), pre(Y,Z),
cat(X,s), cat(Y,np), cat(Z,vp)
    ==>
subj(X,Y), eq(X,Z)
```

a. $S[pss]\backslash NP_i \Rightarrow NP_i\backslash NP_i$
"workers [exposed to it]"

b. $S[adj]\backslash NP_i \Rightarrow NP_i\backslash NP_i$
"a forum [likely to bring attention to the problem]"

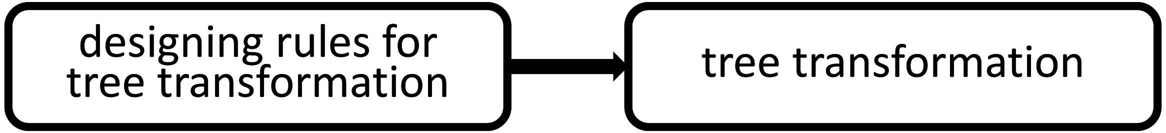
c. $S[ng]\backslash NP_i \Rightarrow NP_i\backslash NP_i$
"signboards [advertising imported cigarettes]"

d. $S[ng]\backslash NP_i \Rightarrow (S\backslash NP_i)\backslash(S\backslash NP_i)$
"become chairman, [succeeding Ian Butler]"

e. $S[dcl]/NP_i \Rightarrow NP_i\backslash NP_i$
"the millions of dollars [it generates]"

Previous Work

- Hand-crafted rules for tree transformation
  (Cahill et al., 2002; Hockenmaier and Steedman, 2007)




```
     s:A
    /   \
np:B     vp:C               dom(A,B), dom(A,C),
 |        |                 dom(C,D), ..
 |        |           =>    pre(B,C),
John     v:D                cat(A,s), cat(C,vp),
          |                 cat(D,v), ..
         left
```

```
dom(X,Y), dom(X,Z), pre(Y,Z),
cat(X,s), cat(Y,np), cat(Z,vp)
    ==>
subj(X,Y), eq(X,Z)
```

a. $S[pss]\backslash NP_i \Rightarrow NP_i\backslash NP_i$
   "workers [exposed to it]"

b. $S[adj]\backslash NP_i \Rightarrow NP_i\backslash NP_i$
   "a forum [likely to bring attention to the problem]"

c. $S[ng]\backslash NP_i \Rightarrow NP_i\backslash NP_i$
   "signboards [advertising imported cigarettes]"

d. $S[ng]\backslash NP_i \Rightarrow (S\backslash NP_i)\backslash(S\backslash NP_i)$
   "become chairman, [succeeding Ian Butler]"

e. $S[dcl]/NP_i \Rightarrow NP_i\backslash NP_i$
   "the millions of dollars [it generates]"

Our Solution – Traditional Classifier

target anno.

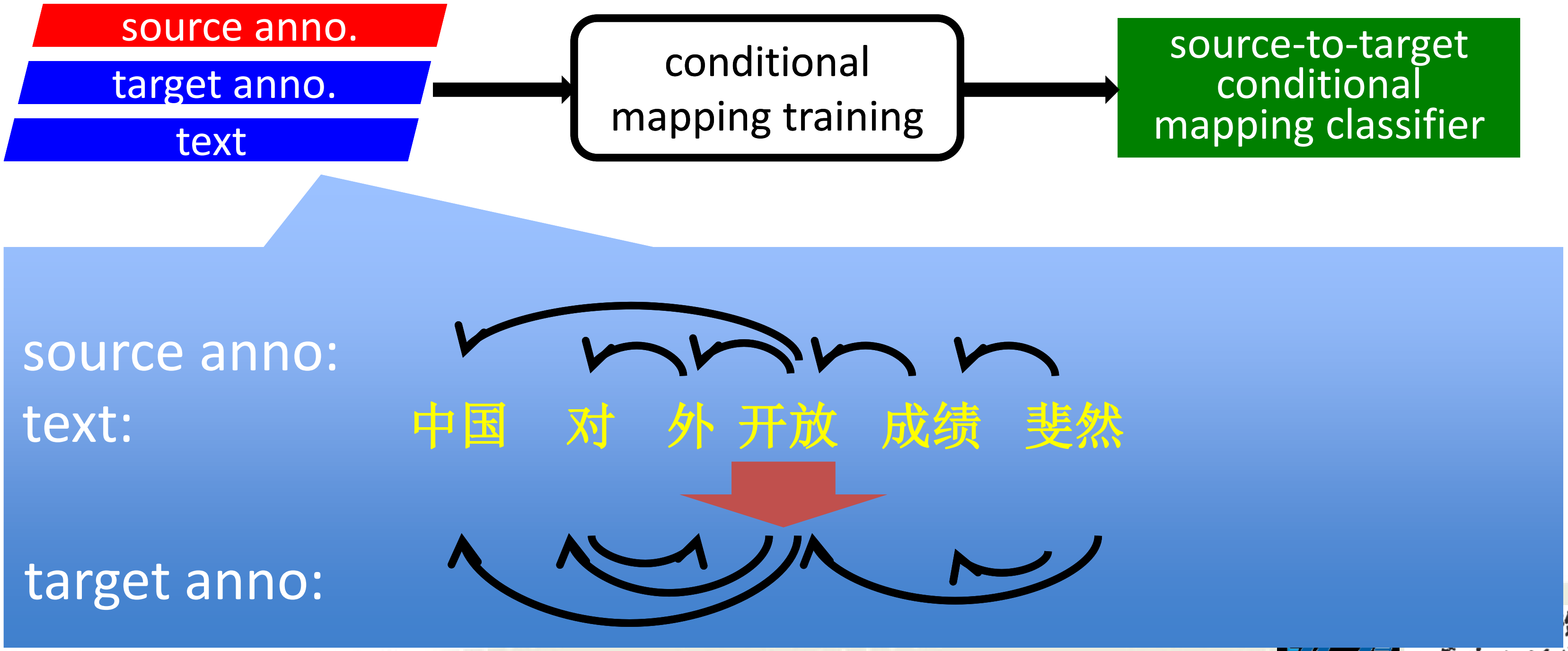
text

→ traditional classifier training → traditional classifier

text: 中国 对 外 开放 成绩 斐然

target anno:

Our Solution



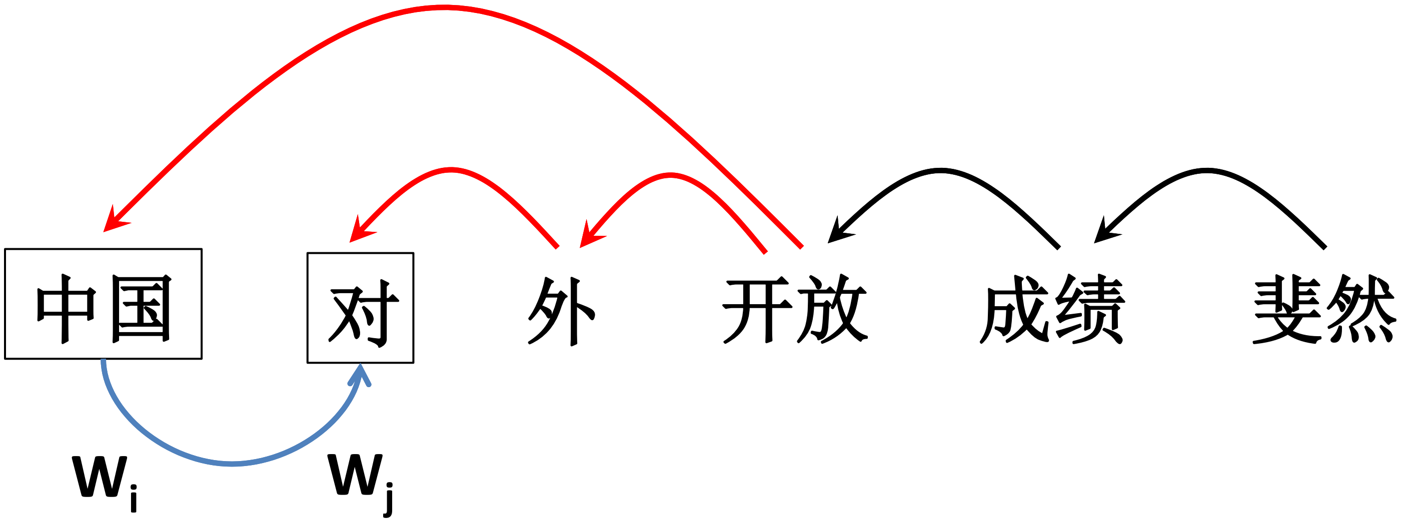
Conditional Mapping Classifier

$\boxed{\text{中国}}$ $\boxed{\text{对}}$ 外 开放 成绩 斐然

W_i W_j

$$P(W_i\text{->}W_j \mid context(i,j))$$

Conditional Mapping Classifier



$$P(W_i \text{->} W_j \mid \text{context}(i,j), \alpha(i,j)=\text{up-down-down})$$

Features – Traditional Classifier Training

Type	Templates	Instances	Type	Templates	Instances
unigram	W_iP_i	W_iP_i=对-P	context	$P_iP_{i+1}P_{j-1}P_j$	$P_iP_{i+1}P_{j-1}P_j$=P-NN-BEG-NR
	W_i	W_i=对		$P_{i-1}P_iP_{j-1}P_j$	$P_{i-1}P_iP_{j-1}P_j$=NR-P-BEG-NR
	P_i	P_i=P		$P_iP_{i+1}P_jP_{j+1}$	$P_iP_{i+1}P_jP_{j+1}$=P-NN-NR-P
	W_jP_j	W_jP_j=中国-NR		$P_{i-1}P_iP_jP_{j+1}$	$P_{i-1}P_iP_jP_{j+1}$=NR-P-NR-P
	W_j	W_j=中国		$P_{i-1}P_iP_{j-1}$	$P_{i-1}P_iP_{j-1}$=NR-P-BEG
	P_j	P_j=NR		$P_{i-1}P_iP_{j+1}$	$P_{i-1}P_iP_{j+1}$=NR-P-P
bigram	$W_iP_iW_jP_j$	$W_iP_iW_jP_j$=对-P-中国-NR		$P_iP_{i+1}P_{j-1}$	$P_iP_{i+1}P_{j-1}$=P-NN-BEG
	$W_iW_jP_j$	$W_iW_jP_j$=对-中国-NR		$P_iP_{i+1}P_{j+1}$	$P_iP_{i+1}P_{j+1}$=NR-P-P
	$P_iW_jP_j$	$P_iW_jP_j$=P-中国-NR		$P_{i-1}P_{j-1}P_j$	$P_{i-1}P_{j-1}P_j$=NR-BEG-NR
	$W_iP_iW_j$	$W_iP_iW_j$=对-P-中国		$P_{i-1}P_jP_{j+1}$	$P_{i-1}P_jP_{j+1}$=NR-NR-P
	$W_iP_iP_j$	$W_iP_iP_j$=对-P-NR		$P_{i+1}P_{j-1}P_j$	$P_{i+1}P_{j-1}P_j$=NN-BEG-NR
	W_iW_j	W_iW_j=对-中国		$P_{i+1}P_jP_{j+1}$	$P_{i+1}P_jP_{j+1}$=NN-NR-P
	W_iP_j	W_iP_j=对-NR		$P_iP_{j-1}P_j$	$P_iP_{j-1}P_j$=P-BEG-NR
	P_iW_j	P_iW_j=P-中国		$P_iP_jP_{j+1}$	$P_iP_jP_{j+1}$=P-NR-P
	P_iP_j	P_iP_j=P-NR		$P_{i-1}P_iP_j$	$P_{i-1}P_iP_j$=NR-P-NR
				$P_iP_{i+1}P_j$	$P_iP_{i+1}P_j$=P-NN-NR

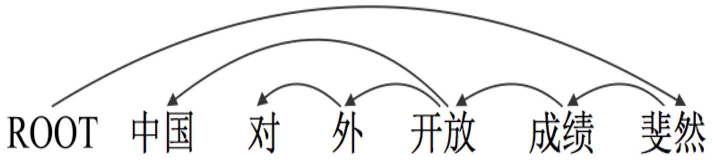
Features – Conditional Mapping Training

Type	Templates	Instances	Type	Templates	Instances
unigram	W_iP_i	W_iP_i=对-P	context	$P_iP_{i+1}P_{j-1}P_j$	$P_iP_{i+1}P_{j-1}P_j$=P-NN-BEG-NR
	W_i	W_i=对		$P_{i-1}P_iP_{j-1}P_j$	$P_{i-1}P_iP_{j-1}P_j$=NR-P-BEG-NR
	P_i	P_i=P		$P_iP_{i+1}P_jP_{j+1}$	$P_iP_{i+1}P_jP_{j+1}$=P-NN-NR-P
	W_jP_j	W_jP_j=中国-NR		$P_{i-1}P_iP_jP_{j+1}$	$P_{i-1}P_iP_jP_{j+1}$=NR-P-NR-P
	W_j	W_j=中国		$P_{i-1}P_iP_{j-1}$	$P_{i-1}P_iP_{j-1}$=NR-P-BEG
	P_j	P_j=NR		$P_{i-1}P_iP_{j+1}$	$P_{i-1}P_iP_{j+1}$=NR-P-P
bigram	$W_iP_iW_jP_j$	$W_iP_iW_jP_j$=对-P-中国-NR		$P_iP_{i+1}P_{j-1}$	$P_iP_{i+1}P_{j-1}$=P-NN-BEG
	$W_iW_jP_j$	$W_iW_jP_j$=对-中国-NR		$P_iP_{i+1}P_{j+1}$	$P_iP_{i+1}P_{j+1}$=NR-P-P
	$P_iW_jP_j$	$P_iW_jP_j$=P-中国-NR		$P_{i-1}P_{j-1}P_j$	$P_{i-1}P_{j-1}P_j$=NR-BEG-NR
	$W_iP_iW_j$	$W_iP_iW_j$=对-P-中国		$P_{i-1}P_jP_{j+1}$	$P_{i-1}P_jP_{j+1}$=NR-NR-P
	$W_iP_iP_j$	$W_iP_iP_j$=对-P-NR		$P_{i+1}P_{j-1}P_j$	$P_{i+1}P_{j-1}P_j$=NN-BEG-NR
	W_iW_j	W_iW_j=对-中国		$P_{i+1}P_jP_{j+1}$	$P_{i+1}P_jP_{j+1}$=NN-NR-P
	W_iP_j	W_iP_j=对-NR		$P_iP_{j-1}P_j$	$P_iP_{j-1}P_j$=P-BEG-NR
	P_iW_j	P_iW_j=P-中国		$P_iP_jP_{j+1}$	$P_iP_jP_{j+1}$=P-NR-P
	P_iP_j	P_iP_j=P-NR		$P_{i-1}P_iP_j$	$P_{i-1}P_iP_j$=NR-P-NR
				$P_iP_{i+1}P_j$	$P_iP_{i+1}P_j$=P-NN-NR

Experiment Setup

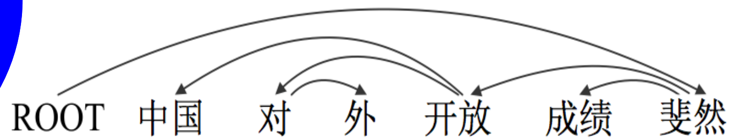
- Target corpus:
  Semantic Dependency Treebank

0.3M words



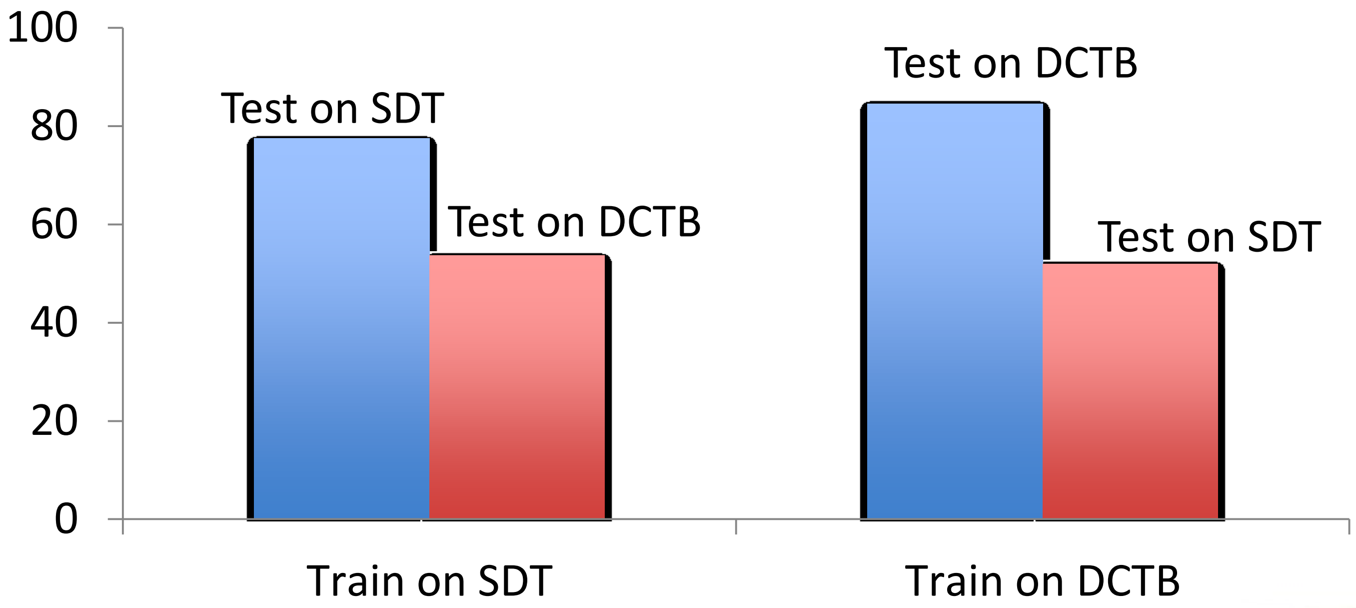
- Source corpus:
  Penn Chinese Treebank 5.0

1M words

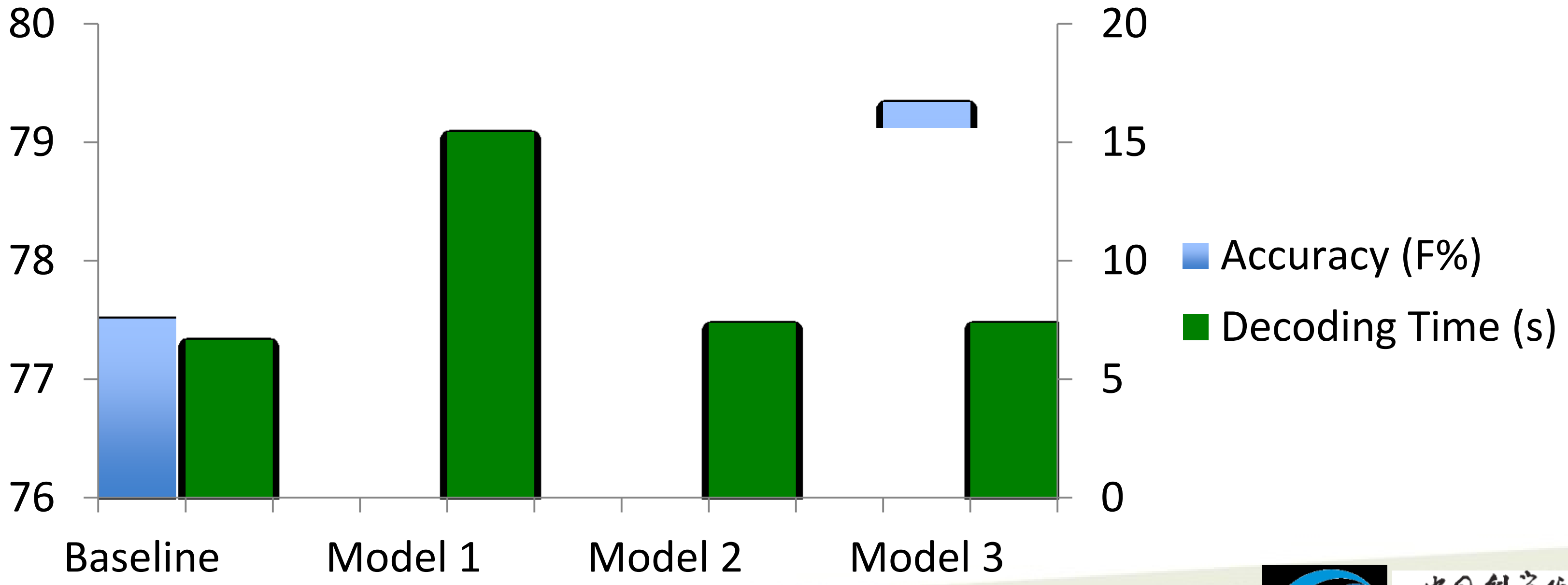


- Classification:
  Averaged perceptron

Baseline Models



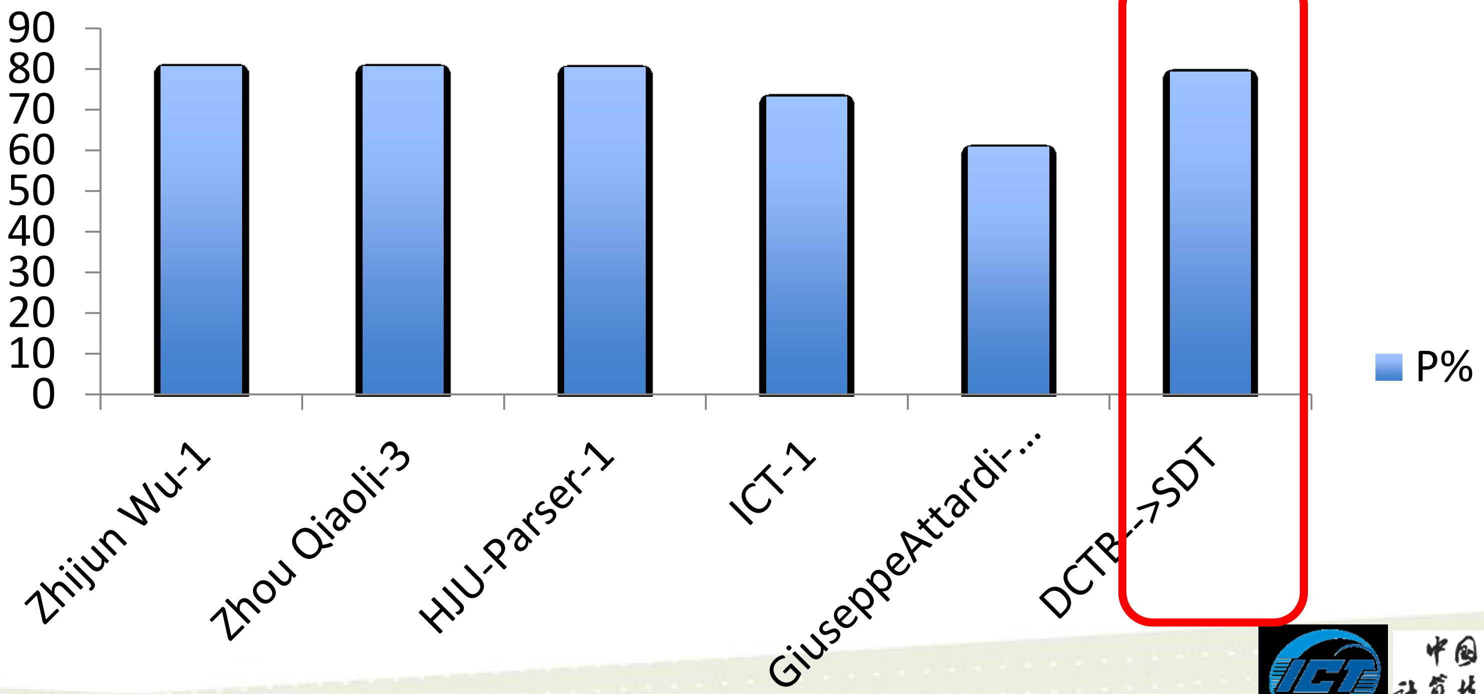
Cross-standard Adaptation for Dependency Parsing



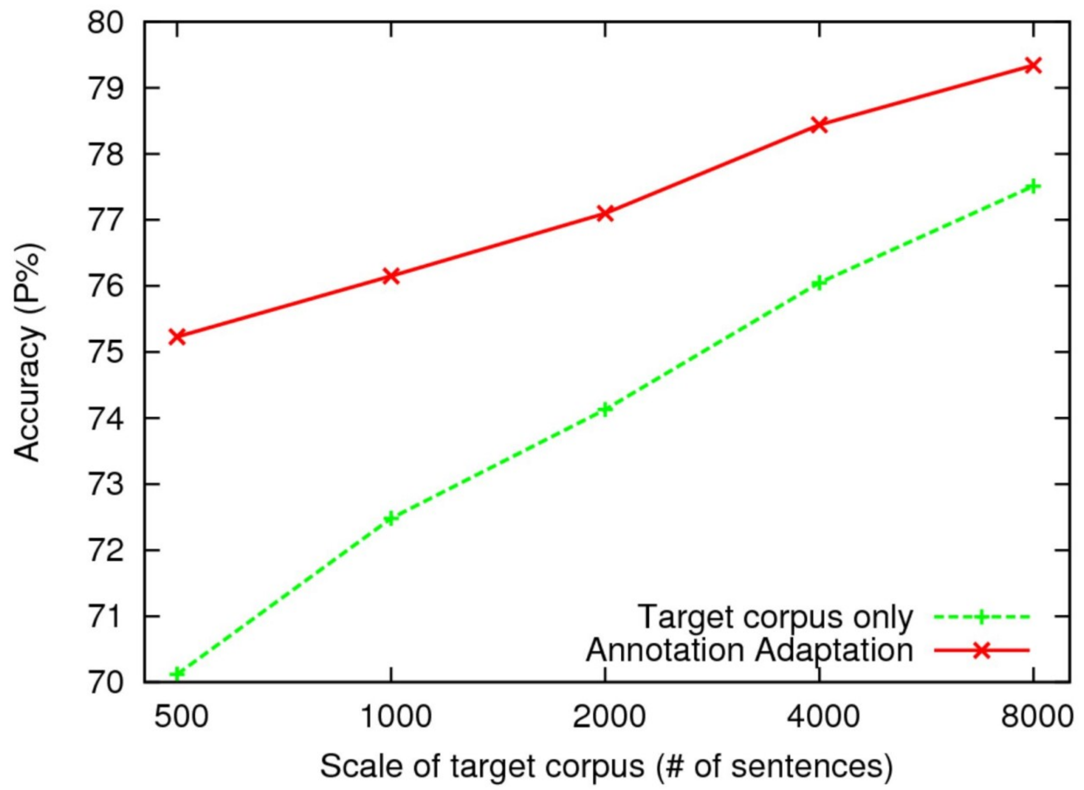
Our Work vs. Non-adaptation Work

System	Model	Features	Adaptation
Zhijun Wu-1	Single	local + non-local	No
Zhou Qiaoli-3	Single	local + non-local	No
HJU-Parser-1	Cascaded	character, non-local, multilevel label	No
Our Work	Single	local	Yes

Our Work vs. Previous Work



Performance wrt #sentence



Our Work vs. Previous Adaptation Work

Representative Previous Work	Automatic/Manual	Method
(Cahill et al. 2002)	Manual	Rule-based Transfer
(Hockenmaier and Steedman 2007)	Manual	Rule-based Transfer
Our Work	Automatic	Statistical

Publications

- Wenbin Jiang, Yajuan Lü, Liang Huang and Qun Liu. 2014. Automatic Adaptation of Annotations. To appear in *Computational Linguistics*.

Outline

Introduction

Cross-Standard Adaptation

Cross-Lingual Adaptation

Experiments on Irish Processing

Conclusion

Outline

- Introduction
- Cross-Standard Adaptation
- **Cross-Lingual Adaptation**
  - **Decomposed Projection**
  - Word Segmentation
  - Dependency Parsing
- Experiments on Irish Processing
- Conclusion

Decomposed Projection

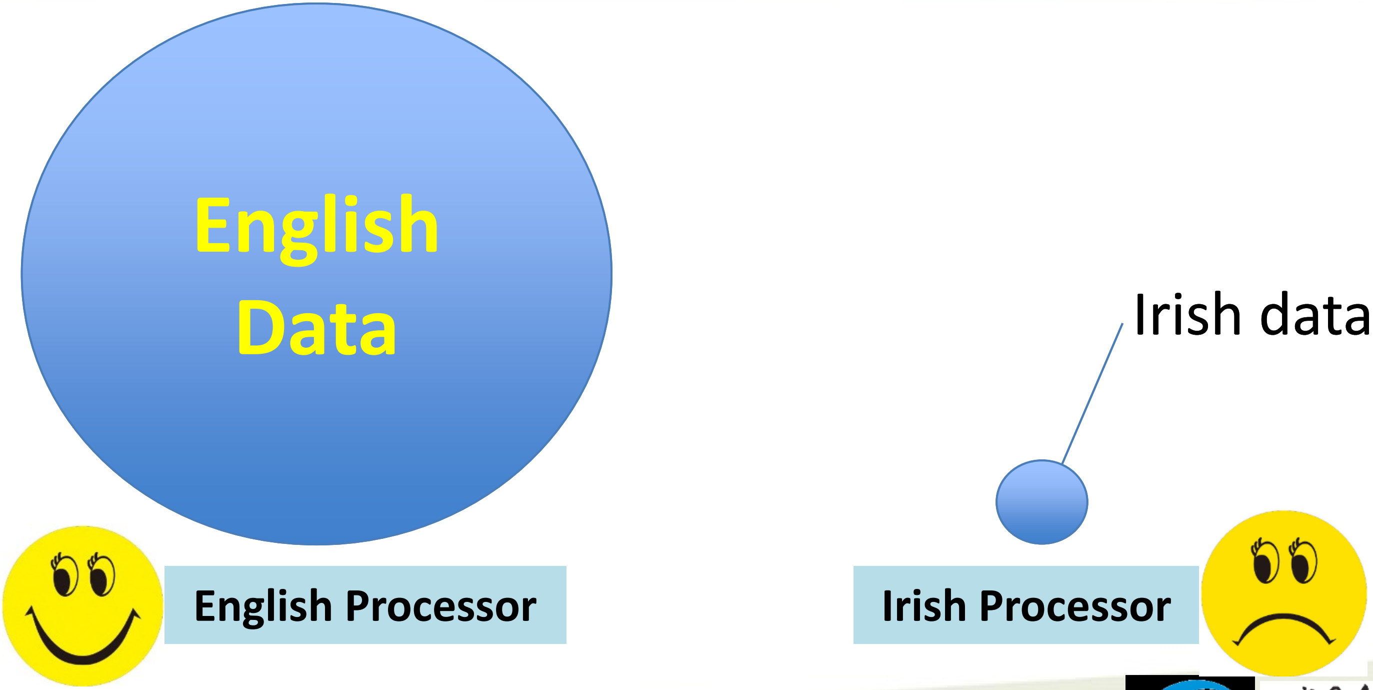
for Cross-lingual Adaptation

Cross-lingual Adaptation

English Data

Cross-lingual Adaptation

Cross-lingual Adaptation



Cross-lingual Adaptation

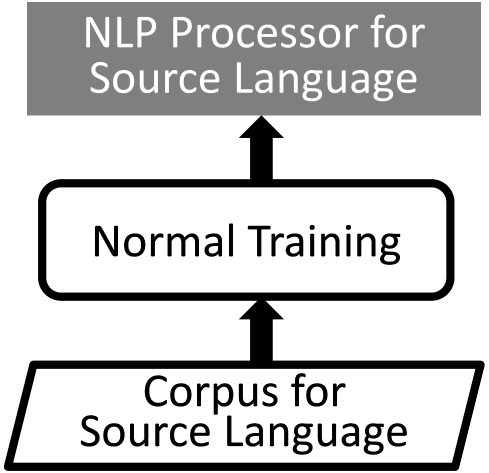
English Data

Irish data

English Processor → Irish Processor

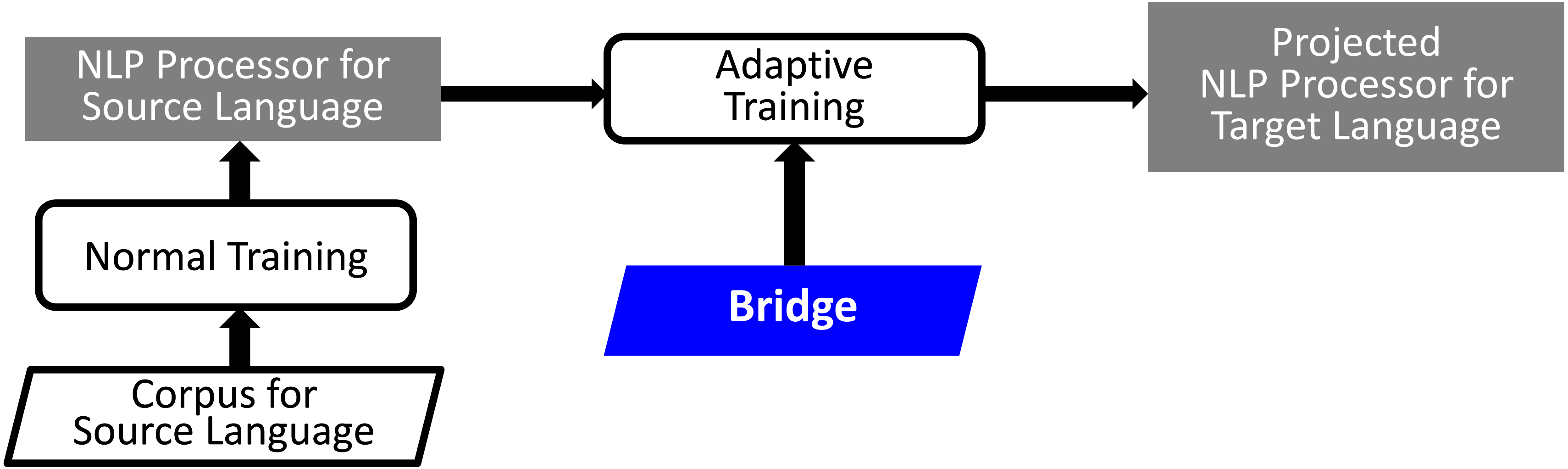
Language Adaptation

Normal Training



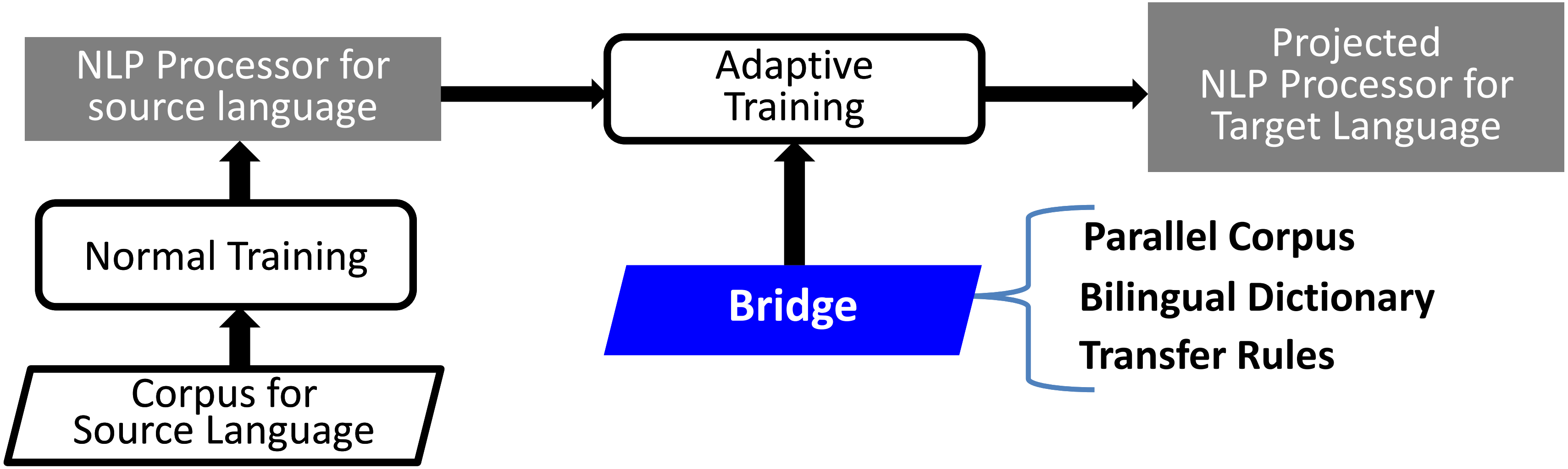
Language Adaptation

Cross-lingual Adaptation Training

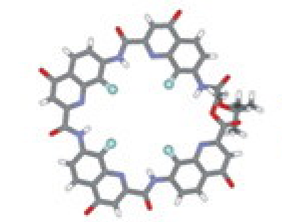


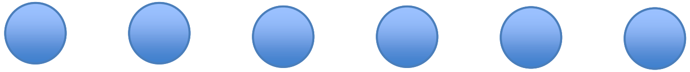
Language Adaptation

Cross-lingual Adaptation Training

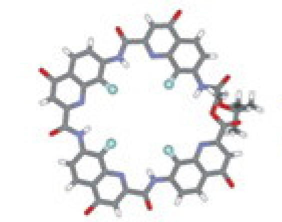


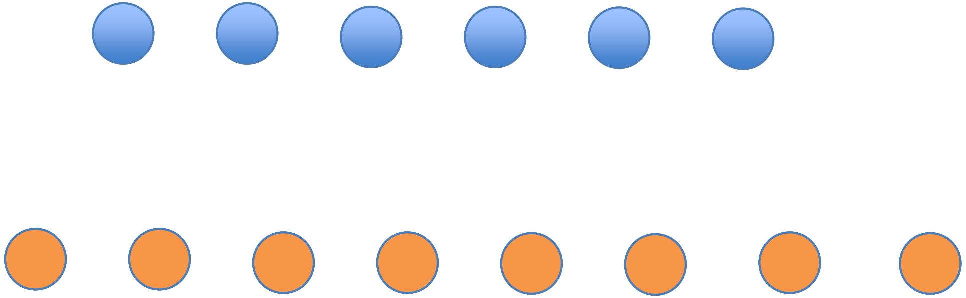
Direct Projection

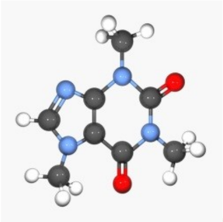




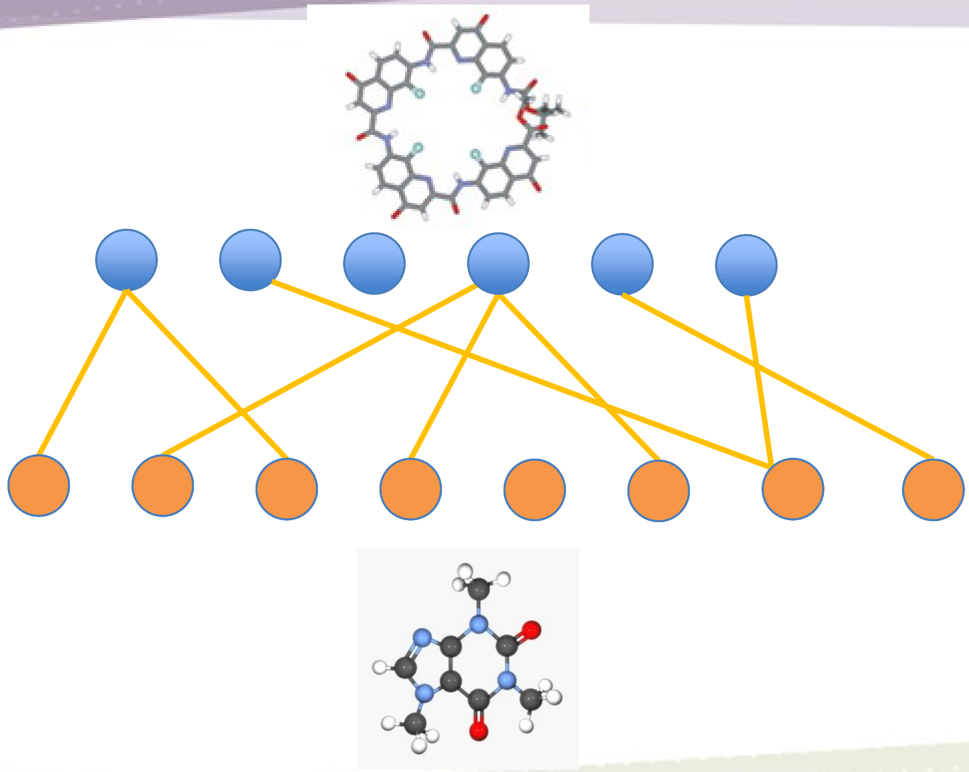
Direct Projection



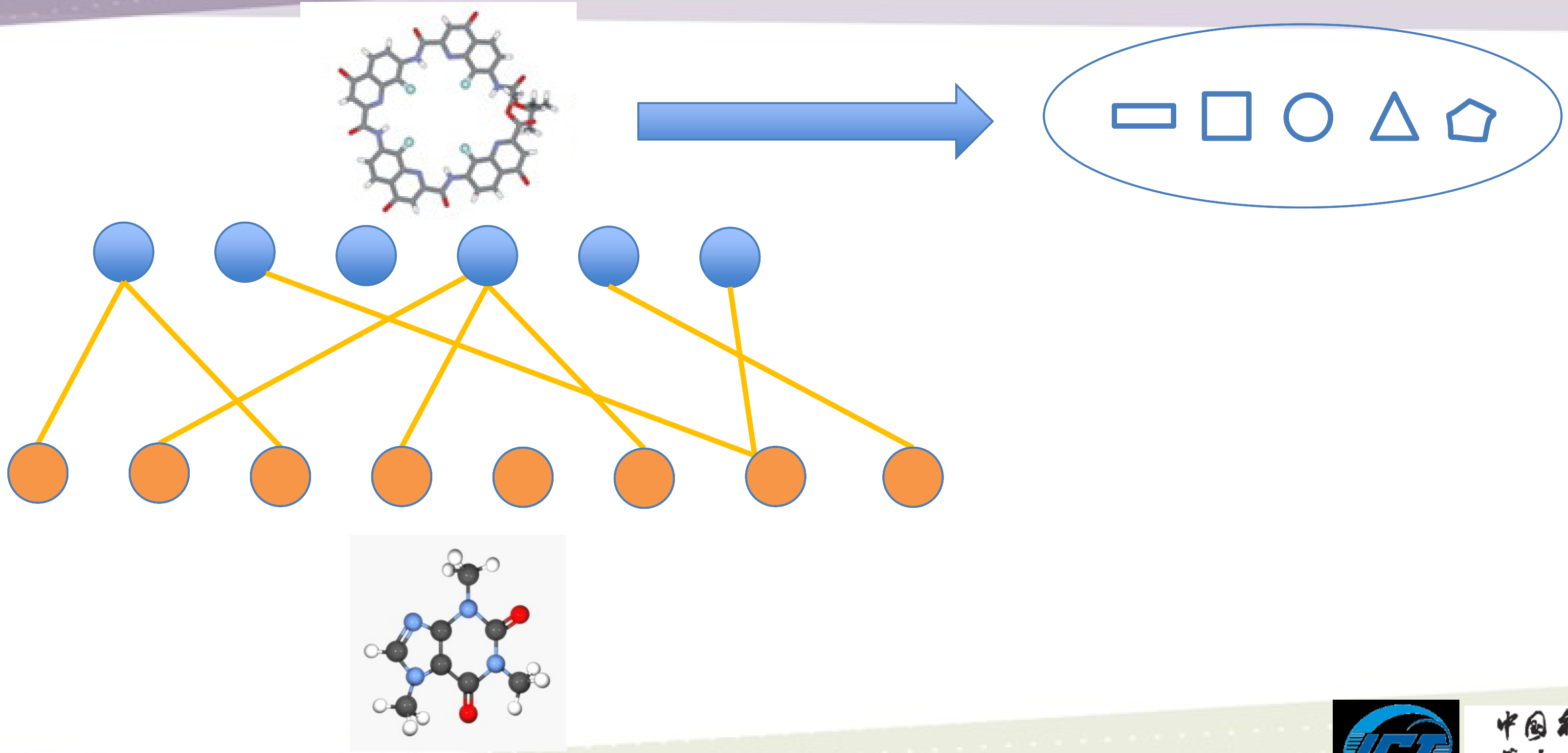




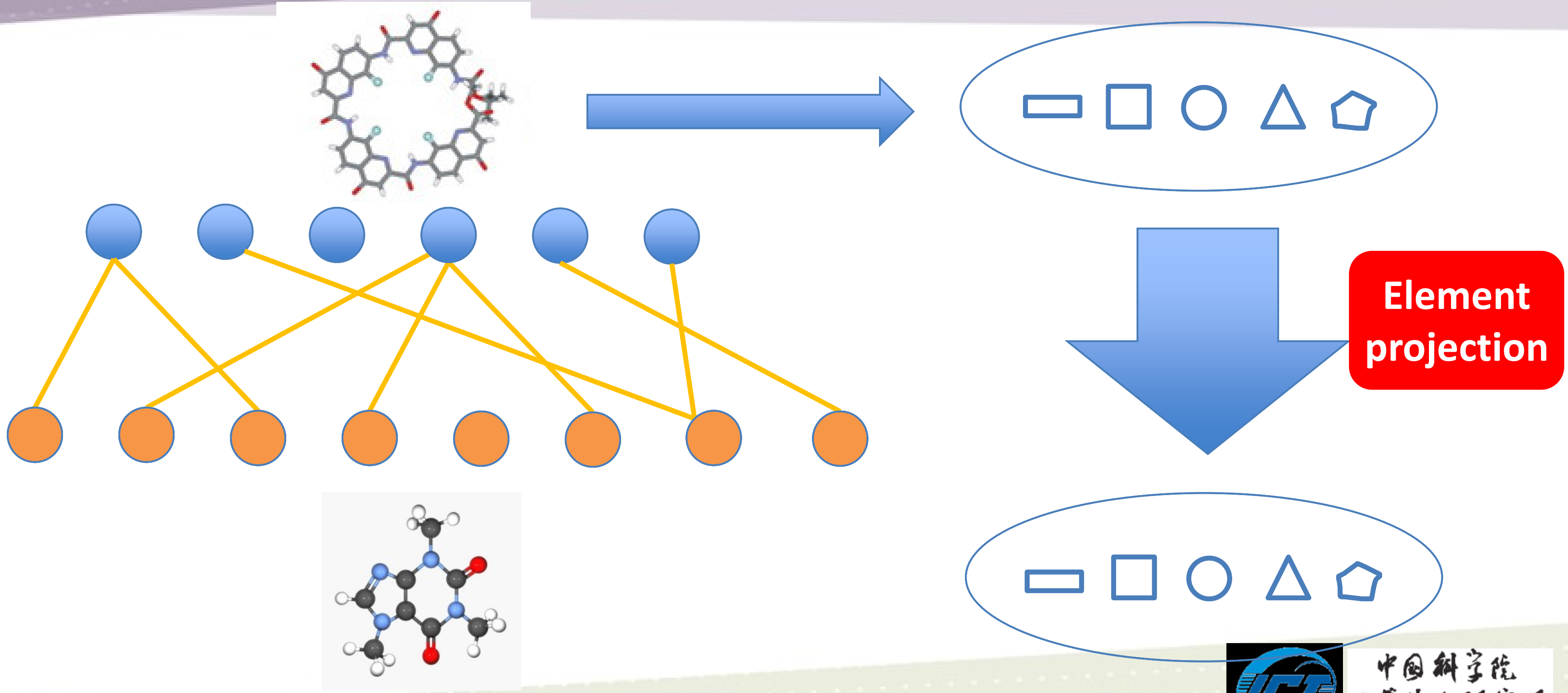
Direct Projection



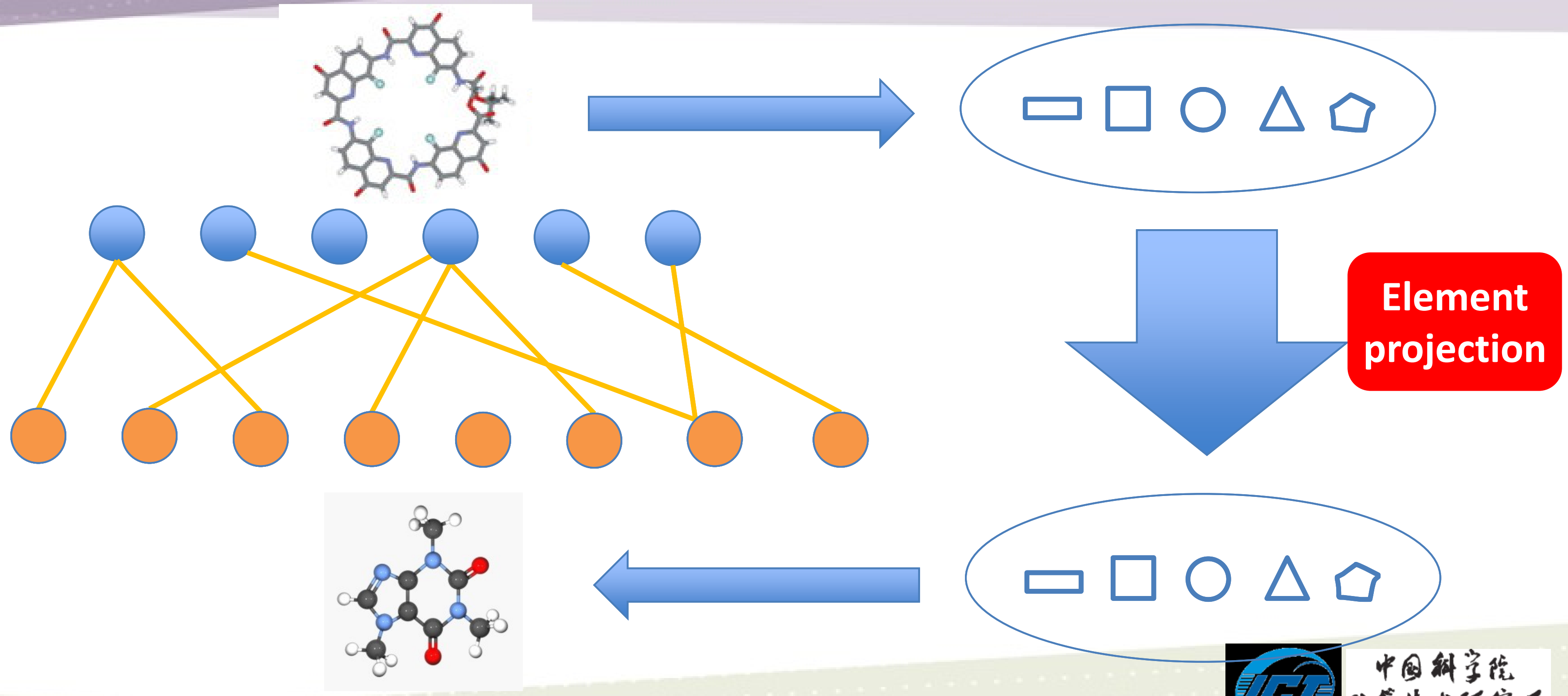
Decomposed Projection



Decomposed Projection



Decomposed Projection



Decomposed Projection

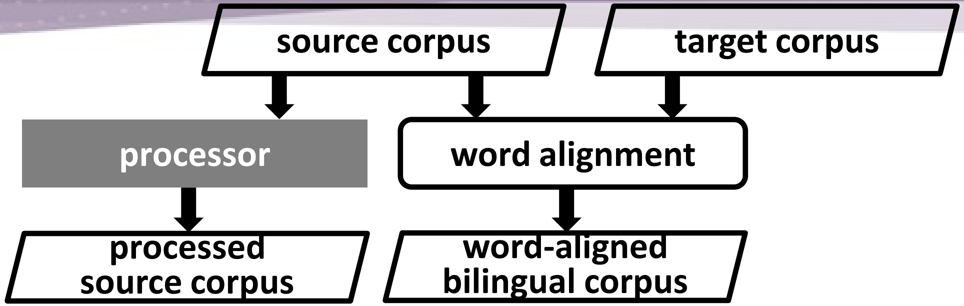
source corpus

target corpus

Input:

source corpus and target corpus correspond to source and target languages

Decomposed Projection

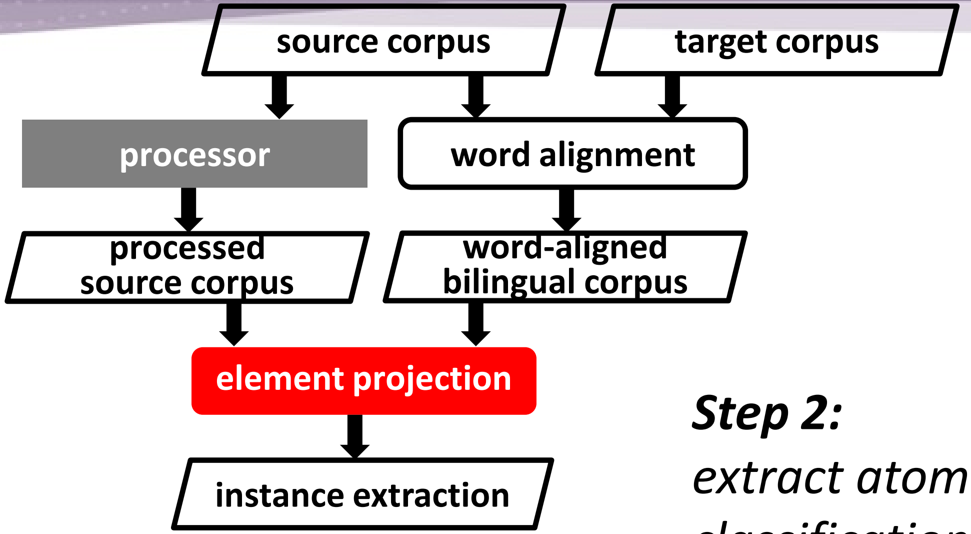


Step 1:

- process the source corpus with the existing NLP processor

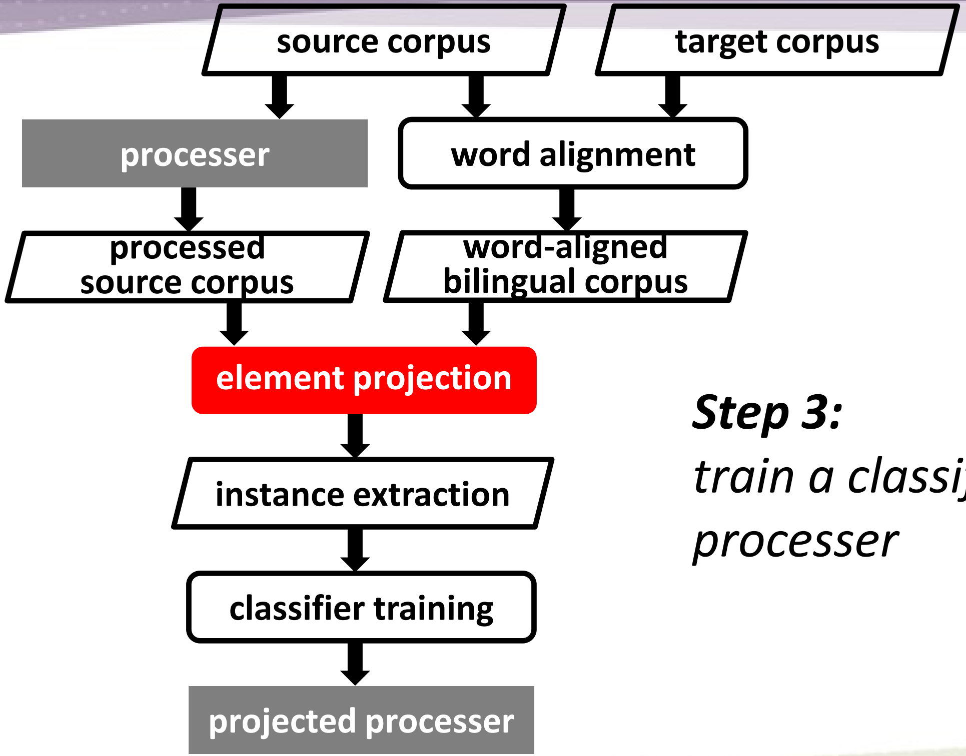
- perform word alignment between source and target corpora

Decomposed Projection



Step 2:
extract atomic instances, e.g. character classification instances for word segmentation, and word-pair dependency instances for dependency parsing

Decomposed Projection



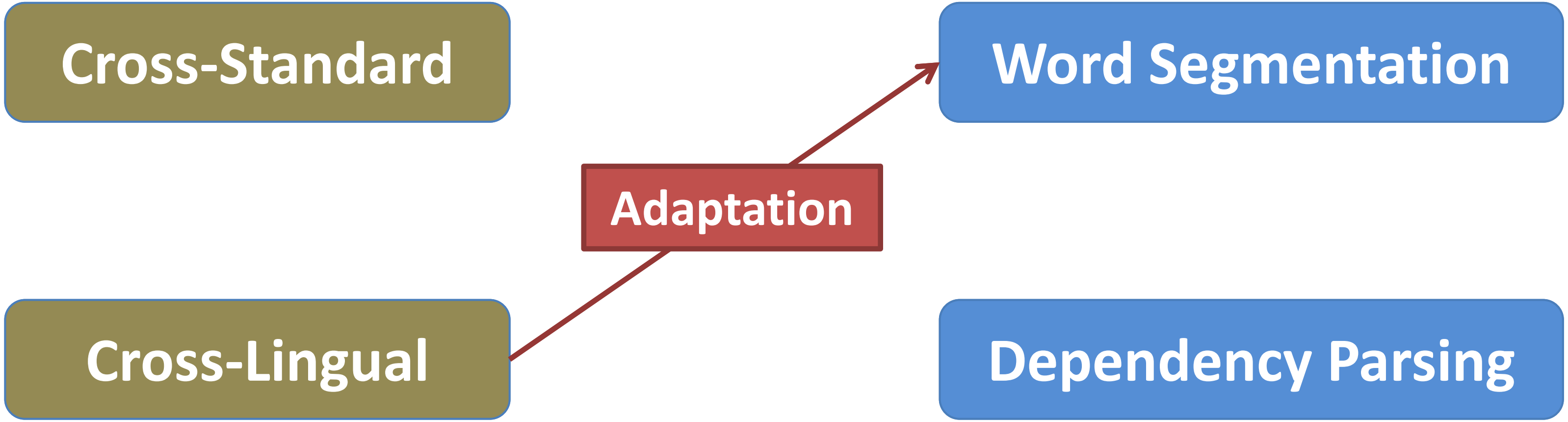
Step 3:

train a classifier which is the final projected processer

Outline

- Introduction
- Cross-Standard Adaptation
- **Cross-Lingual Adaptation**
  - Decomposed Projection
  - **Word Segmentation**
  - Dependency Parsing
- Experiments on Irish Processing
- Conclusion

Problem

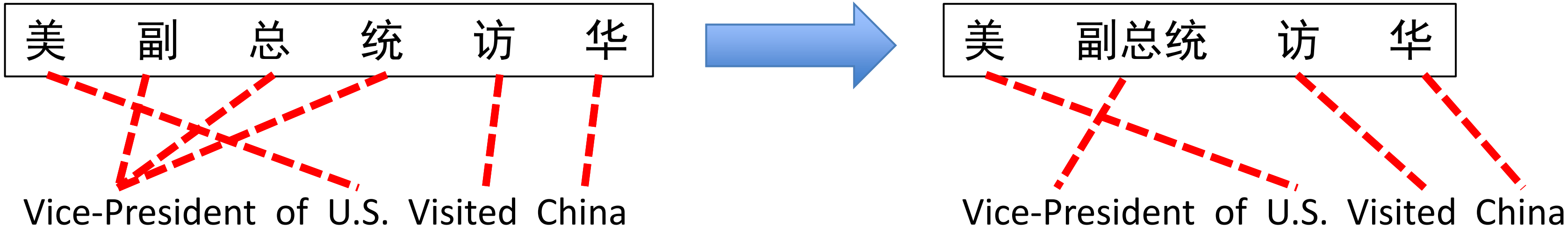

Cross-Standard
Word Segmentation
Adaptation
Cross-Lingual
Dependency Parsing

Cross-lingual Adaptation for Word Segmentation

- English is naturally segmented
- Can we use word boundary information from English text to learn a Chinese segmentation algorithm, by using an English-Chinese bilingual corpus as a bridge?

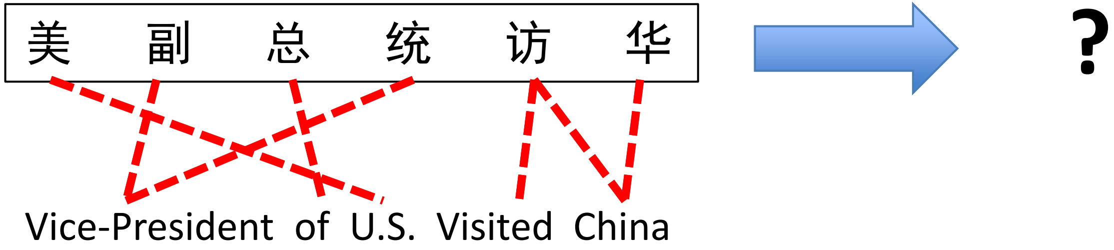
Cross-lingual Adaptation for Word Segmentation

- Cross-lingual adaptation for word segmentation aims to learn or improve a word segmenter resorting to bitext aligned to a language with natural word boundaries (or segmented)



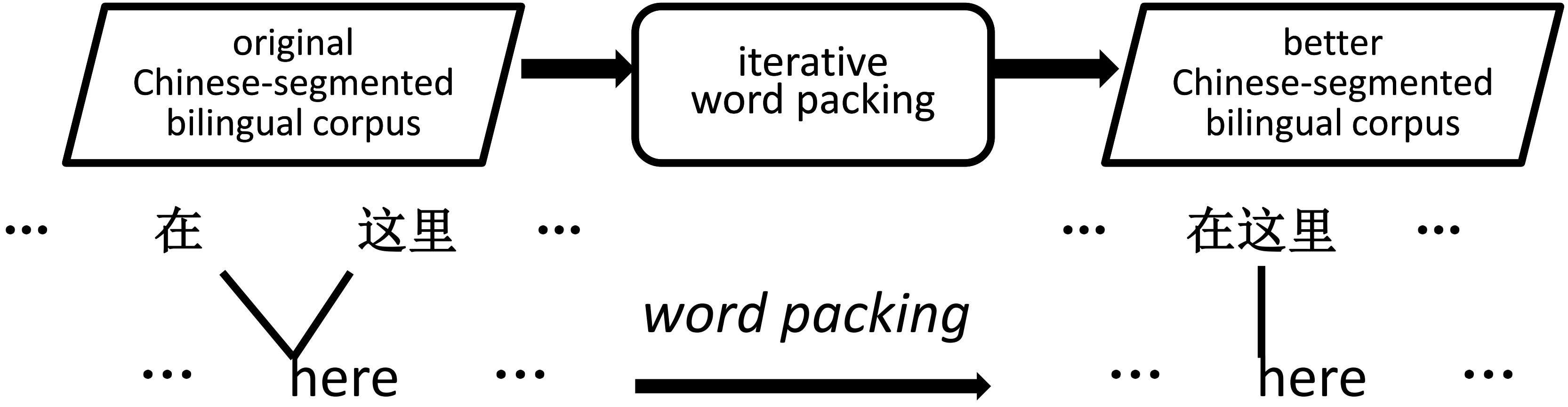
Cross-lingual Adaptation for Word Segmentation

- It is not always possible to project an English sentence to a Chinese word segmentation because of the noisy word alignments:



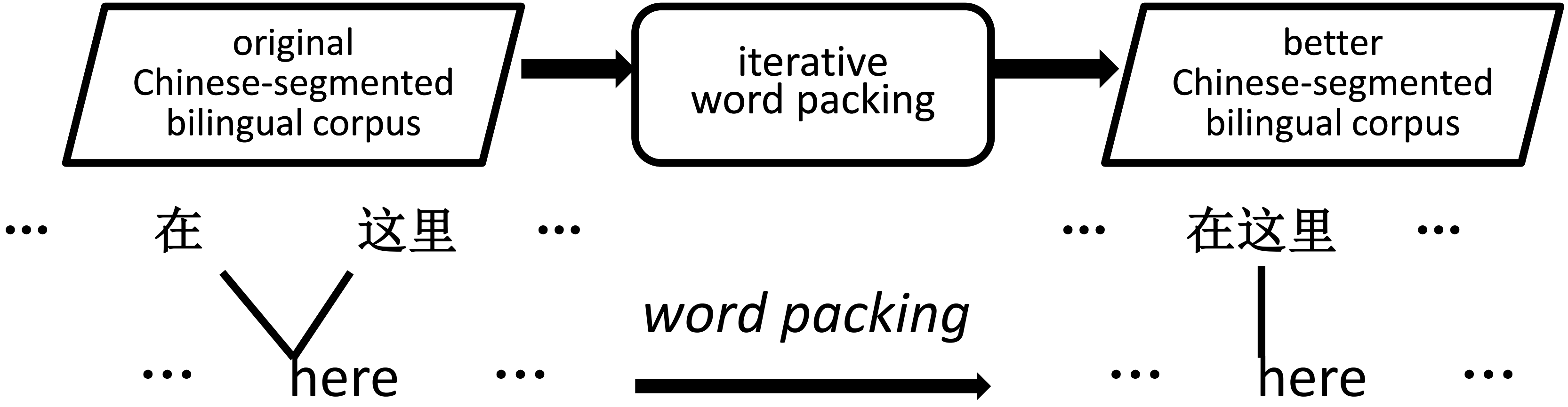
Previous Work

- Bilingually optimized word segmentation by word packing (Ma and Way, 2007)



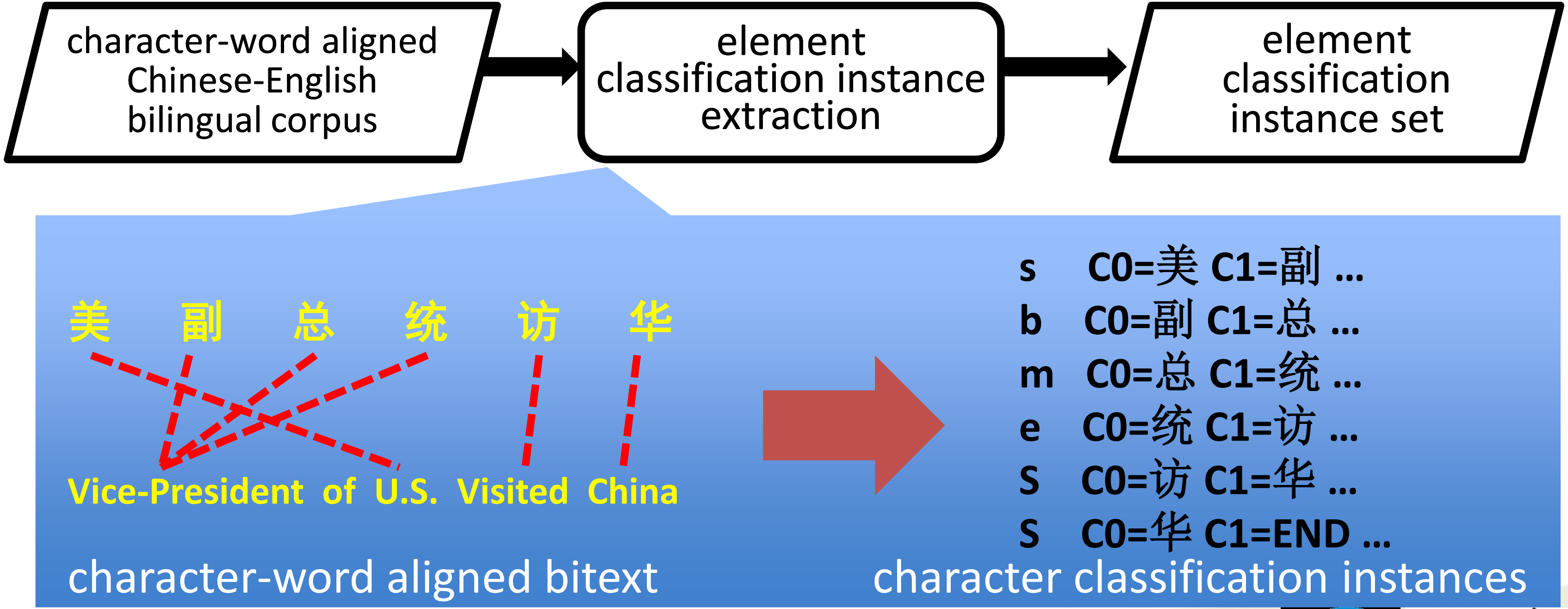
Previous Work

- Bilingually optimized word segmentation by word packing (Ma and Way, 2007)



Structure Projection

Our Solution



Our Solution

character-word aligned Chinese-English bilingual corpus → element classification instance extraction → element classification instance set

美 副 总 统 访 华

Vice-President of U.S. Visited China

character-word aligned bitext

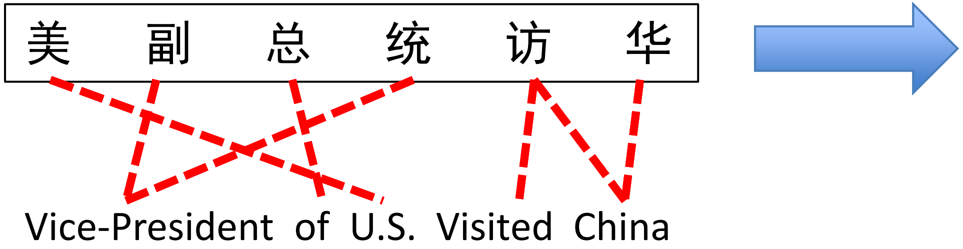
s C0=美 C1=副 …
b C0=副 C1=总 …
m C0=总 C1=统 …
e C0=统 C1=访 …
S C0=访 C1=华 …
S C0=华 C1=END …

character classification instances

Decomposed Projection

Instance Extraction Criterion

- Only when:
  - A English word is aligned to several adjacent Chinese characters
  - None of these Chinese characters is aligned to other English word
- Then these Chinese characters can be extracted as training instances for the training of Chinese word segmentation



Only 美 and 总 can be extracted as instances

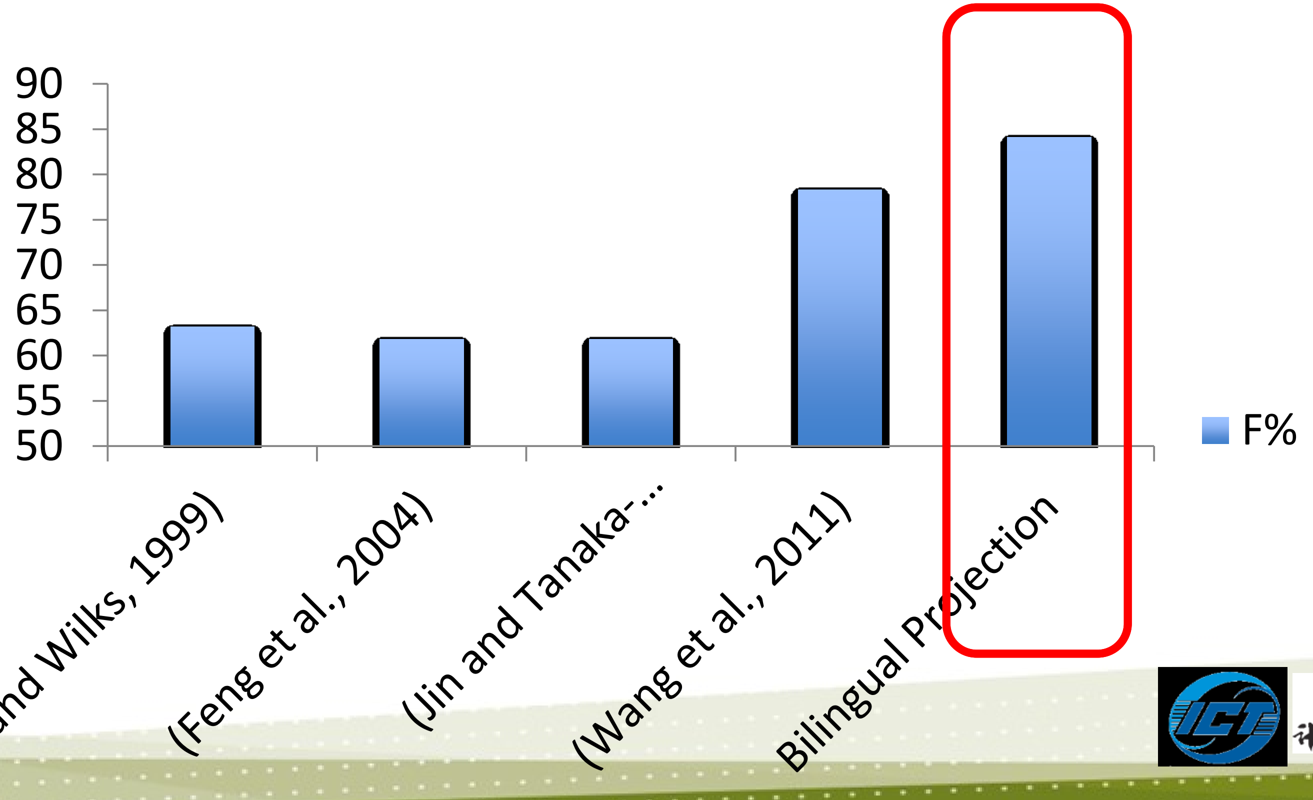
Decomposed Projection for Word Segmentation

Structure	**Word Sequence**	**Vice-President of U.S.=>美 副总统**
Element	**Character + Boundary Annotation**	**Vice-President of U.S.=>美 副 总 统** **S B M E**

Experiment

- Training Data:
  Bilingual corpus: FBIS Chinese-English Corpus
  - # of Chinese words: 6.9M
  - # of English words: 8.9M
  - # of sentence pairs: 239K

Our Work vs. Unsupervised Work



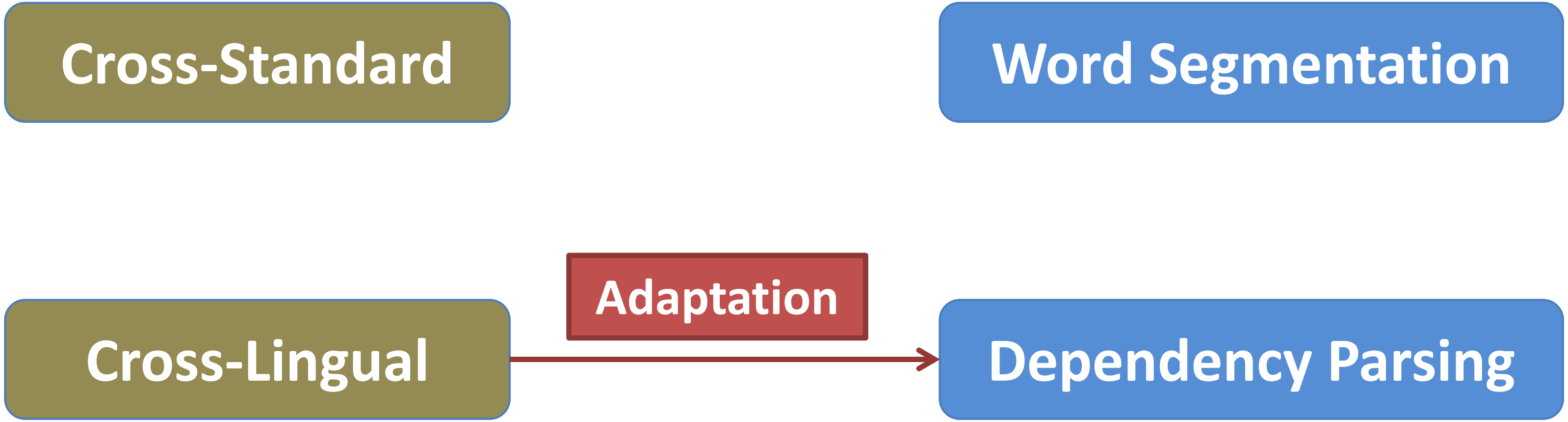
Comparison with Previous Adaptation Work

Representative Previous Work	Method	Language Similarity Requirement	Alignment Error Tolerance
(Ma and Way, 2007)	Structure Projection	Low	Low
Our Work	Decomposed Projection	Low	High

Outline

- Introduction
- Cross-Standard Adaptation
- **Cross-Lingual Adaptation**
  - Decomposed Projection
  - Word Segmentation
  - **Dependency Parsing**
- Experiments on Irish Processing
- Conclusion

Problem


Cross-Standard
Word Segmentation
Adaptation
Cross-Lingual
Dependency Parsing

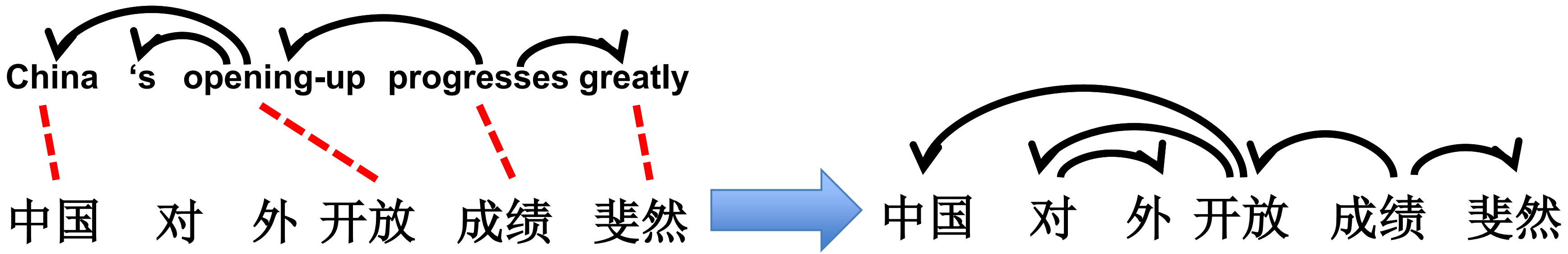
Cross-lingual Adaptation for Dependency Parsing

- English parsing achieves good performance

- For many languages, there is no manually annotated corpus, or the size is very small, however usually there are comparatively large-sized bilingual corpora with English

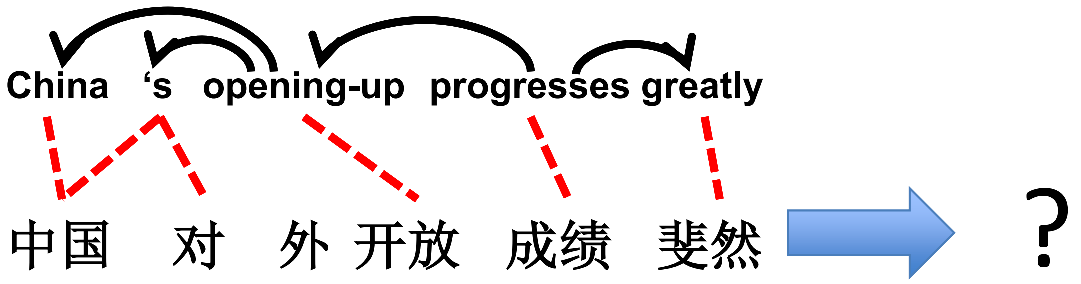
Cross-lingual Adaptation for Dependency Parsing

- Cross-lingual adaptation for dependency parsing aims to learn or improve a dependency parser resorting to bitext aligned to a language with better parsers



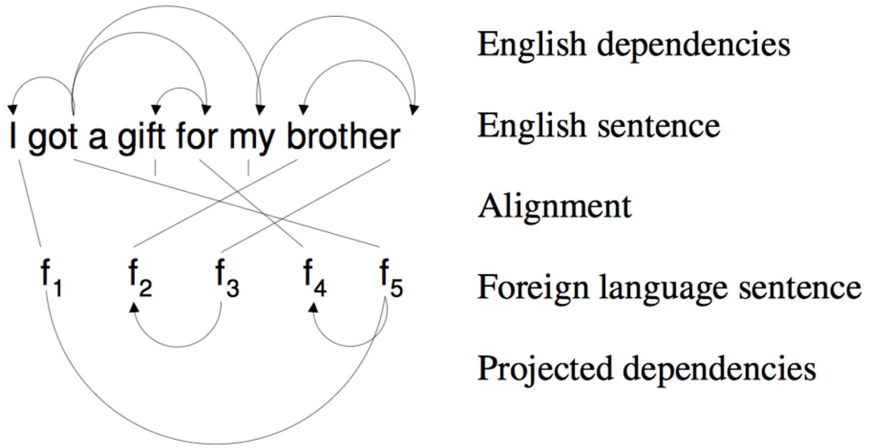
Cross-lingual Adaptation for Dependency Parsing

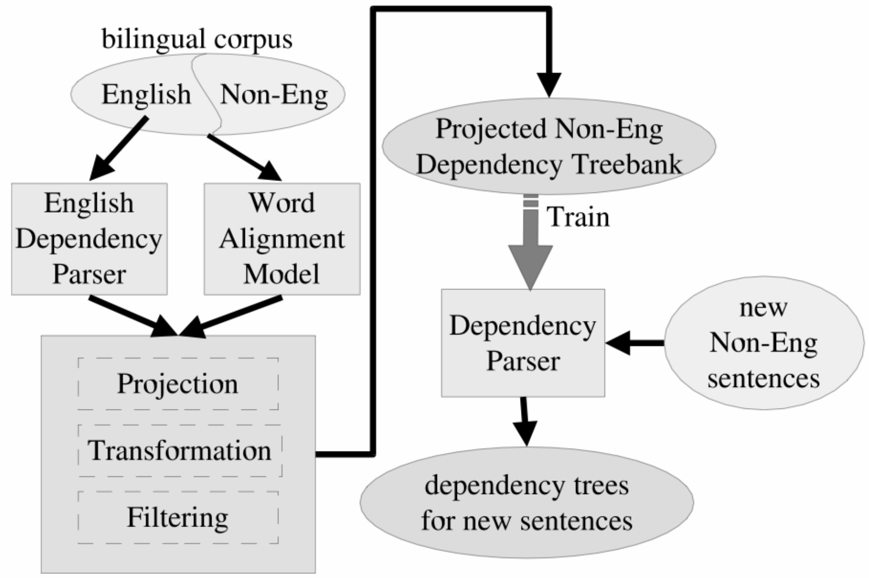
- It is not always possible to project an English dependency tree to a Chinese dependency tree because of the noisy word alignment



Previous Work

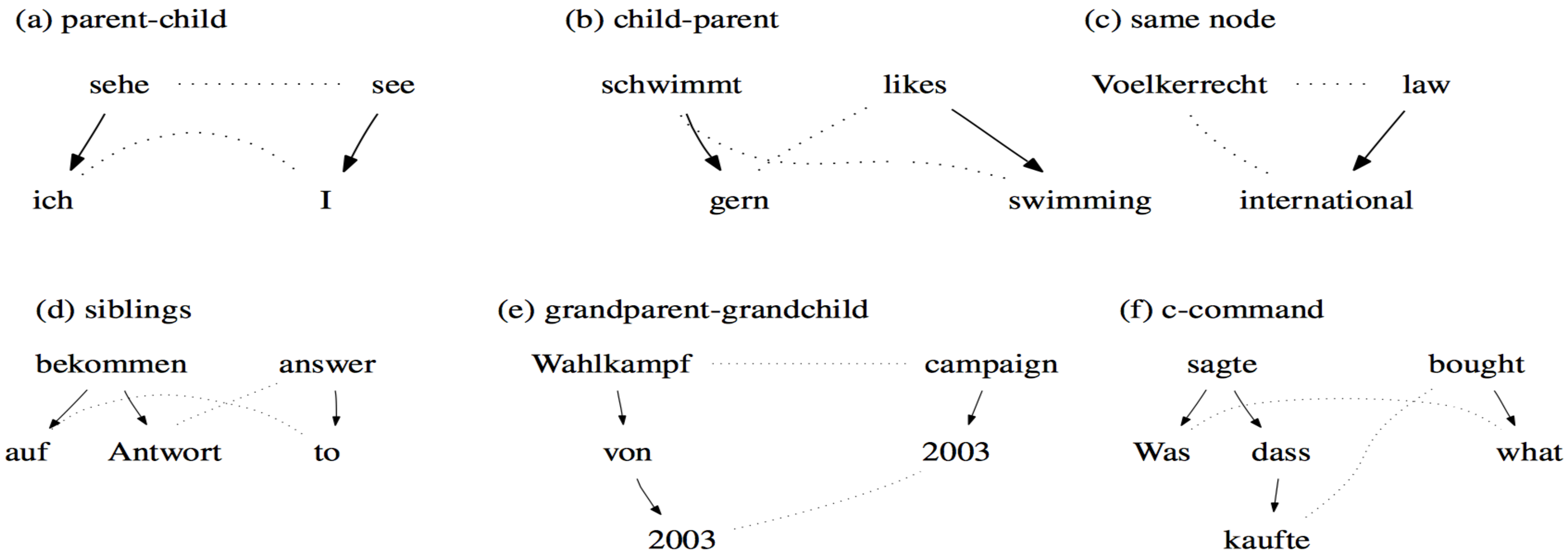
- Direct projection of dependency structures (Hwa et al., 2005)





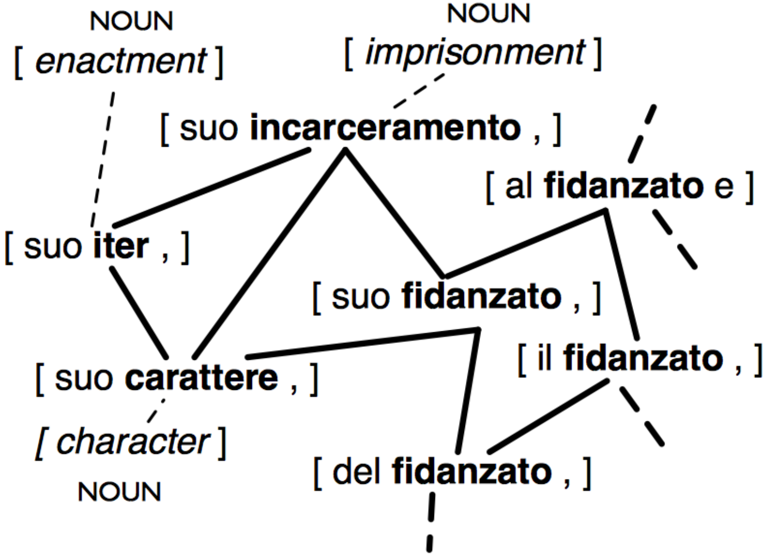
Previous Work

- Optimized projection for dependency with quasi-synchronous grammar (Smith and Eisner, 2009)



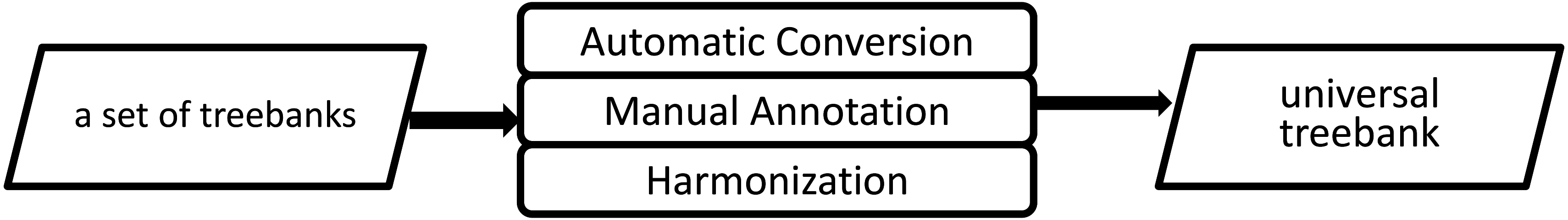
Previous Work

- Optimized projection for POS with graph propagation (Das and Petrov, 2011)



Existing Work

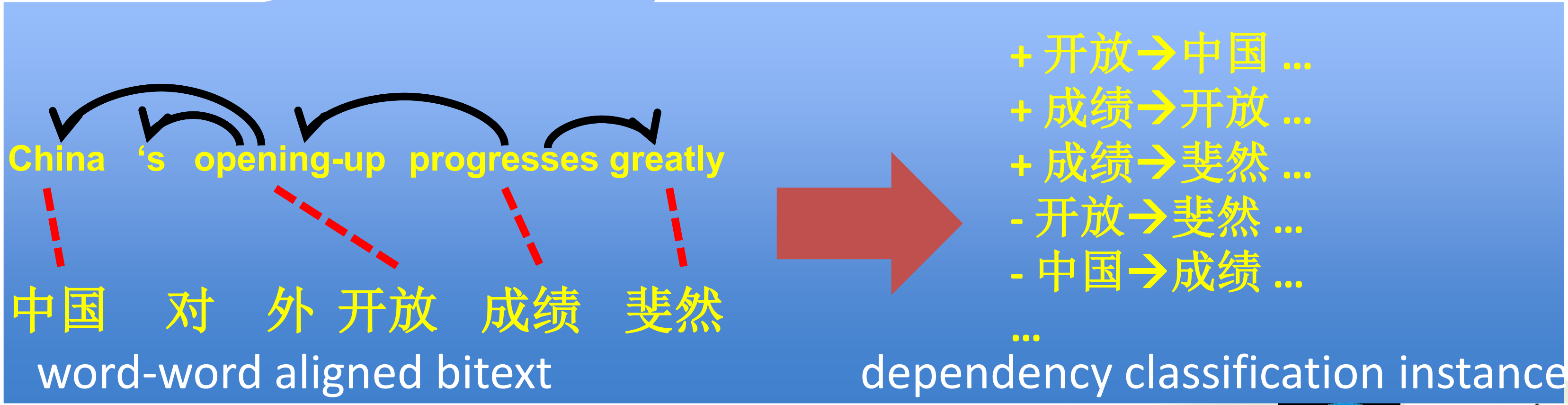
- A collection of universal dependency treebank covering 6 languages (McDonald et al., 2013)



German | English | Swedish | Spanish | French | Korean

Our Solution

character-word aligned Chinese-English bilingual corpus → element classification instance extraction → element classification instance set

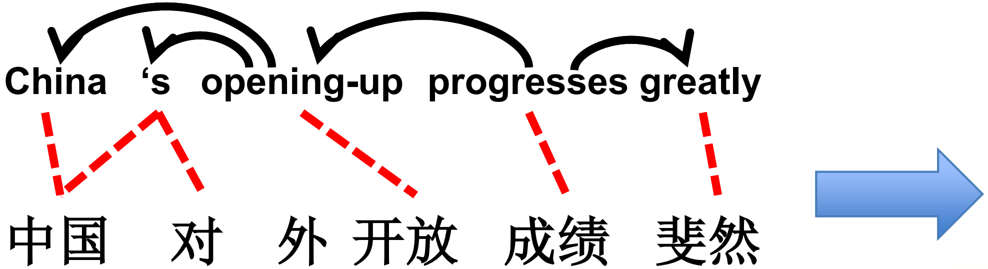


Structure Mapping vs. Decomposed Projection

Structure	**Dependency Tree**	中国 对 外 开放 成绩 斐然
Elements	**Word Pairs with Edges**	成绩 斐然 中国 开放 ~~中国 成绩~~ ~~中国 斐然~~ 对 开放 ~~外 斐然~~

Instance Extraction Criterion

- Only when:
  - A dependency exists between two English words E1 and E2;
  - There are one-to-one alignment between E1←→C1 and E2←→C2;
- Then
  - we can extract C1 and C2 as a instance for Chinese parser training

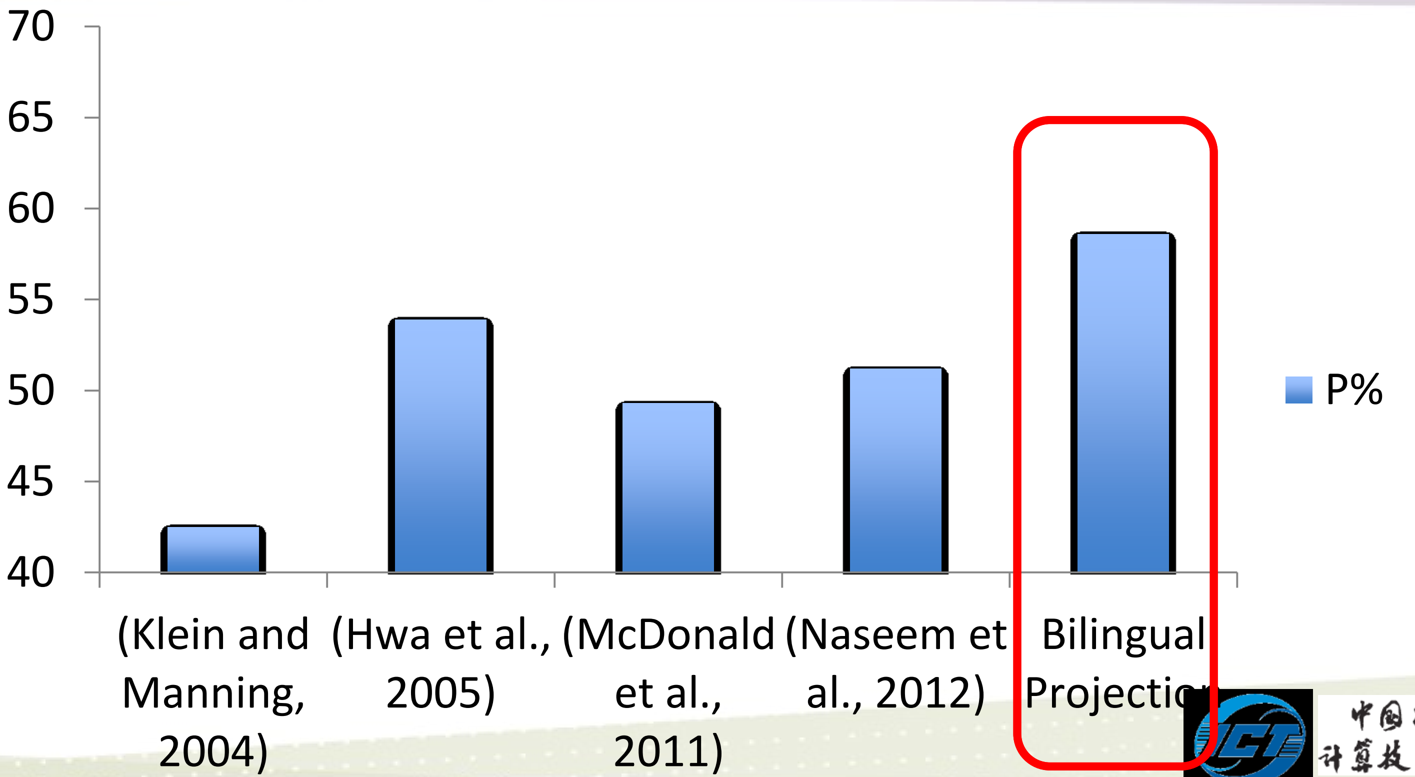


~~? 开放→中国 …~~
+ 成绩→开放 …
+ 成绩→斐然 …
- 开放→斐然 …
~~? 中国→成绩 …~~
…

Experiment

- Training Data:
  Bilingual corpus: FBIS Chinese-English Corpus
  – # of Chinese words: 6.9M
  – # of English words: 8.9M
  – # of sentence pairs: 239K

Experimental Results



Further Improvement

Source Parser

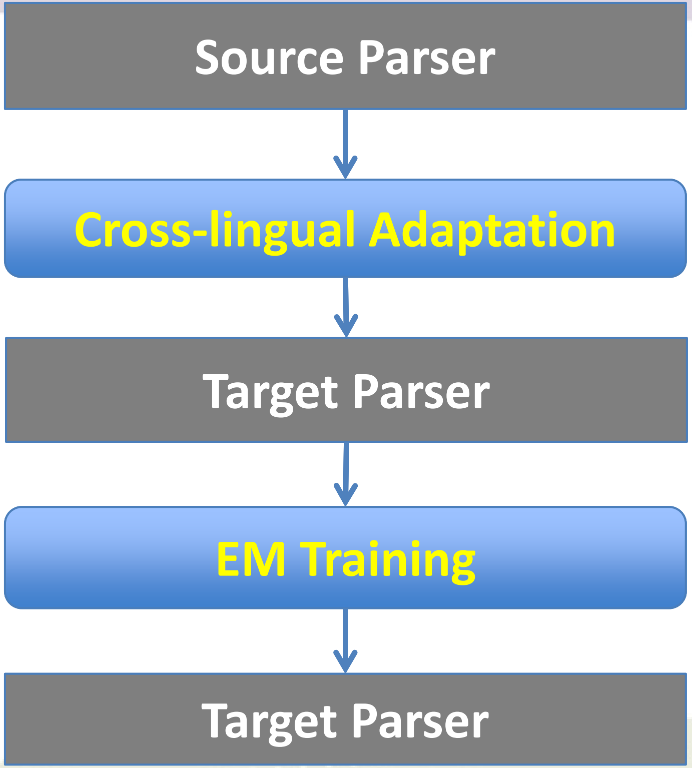
↓

Cross-lingual Adaptation

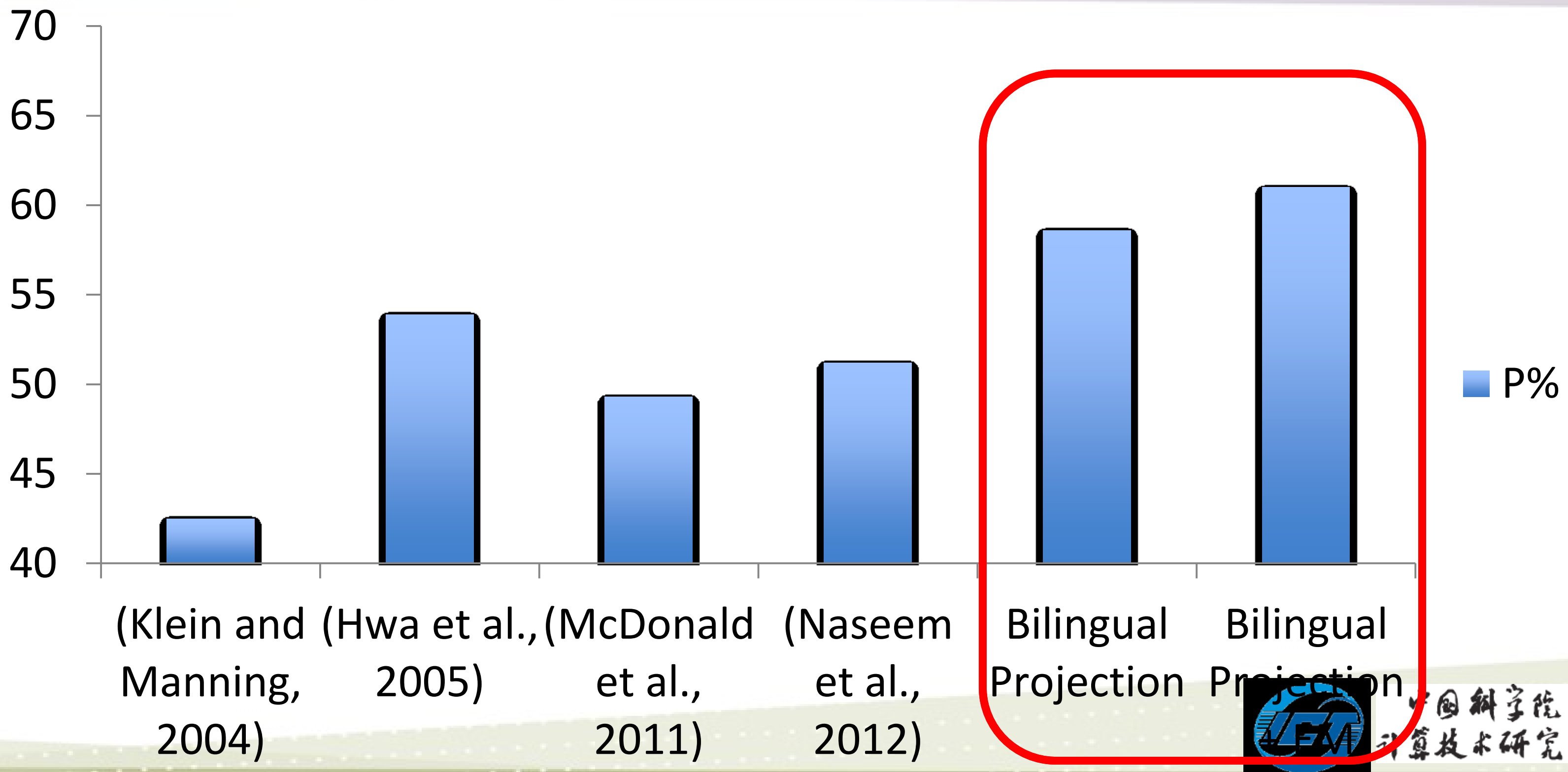
↓

Target Parser

Further Improvement



Experimental Results



Our Work vs. Non-adaptation Work

Representative Previous Work	Method	Time Cost	Annotation Requirement
(Klein and Manning, 2004)	Unsupervised	High	No
(McDonald et al., 2011)	Delexicalized Multi-source Transfer	Low	No
(McDonald et al. 2013)	Universal Grammar	Low	Yes
Our Work	Decomposed Projection	Low	No

Our Work vs. Previous Adaptation Work

Representative Previous Work	Method	Language Similarity Requirement	Alignment Error Tolerance
(Hwa et al., 2005)	Direct correspondence assumption	High	Low
(Smith and Eisner, 2009)	Quasi-synchronous Grammar	Low	High
(Das and Petrov, 2011)	Graph propagation	Low	Low
Our Work	Decomposed Projection	Low	High

Publications

- Wenbin Jiang and Qun Liu. 2010. Dependency Parsing and Projection Based on Word-Pair Classification. In *Proceedings of ACL 2010*, Uppsala, Sweden.

- Kai Liu, Yajuan Lü, Wenbin Jiang and Qun Liu. 2013. Bilingually-Guided Monolingual Dependency Grammar Induction. In *Proceedings of ACL 2013*, Sofia, Bulgaria.

Outline

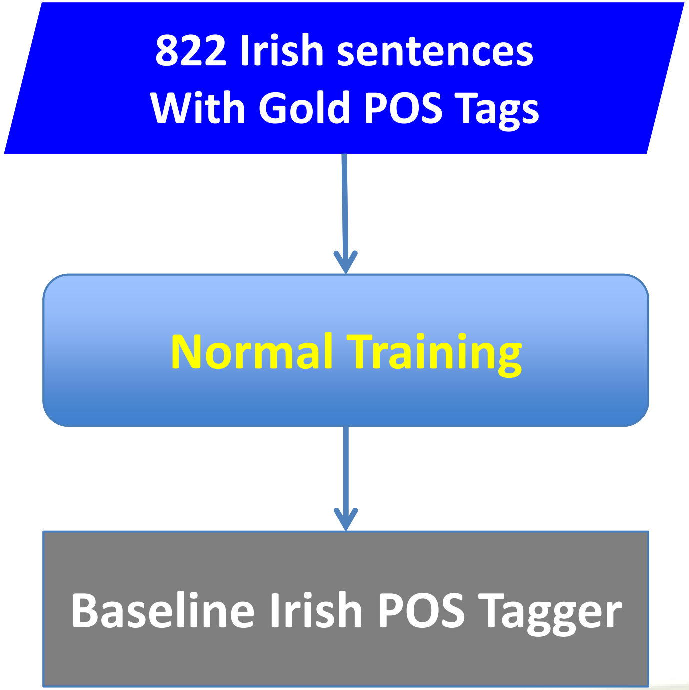
- Introduction
- Cross-Standard Adaptation
- Cross-Lingual Adaptation
- **Experiments on Irish Processing**
- Conclusion

Data

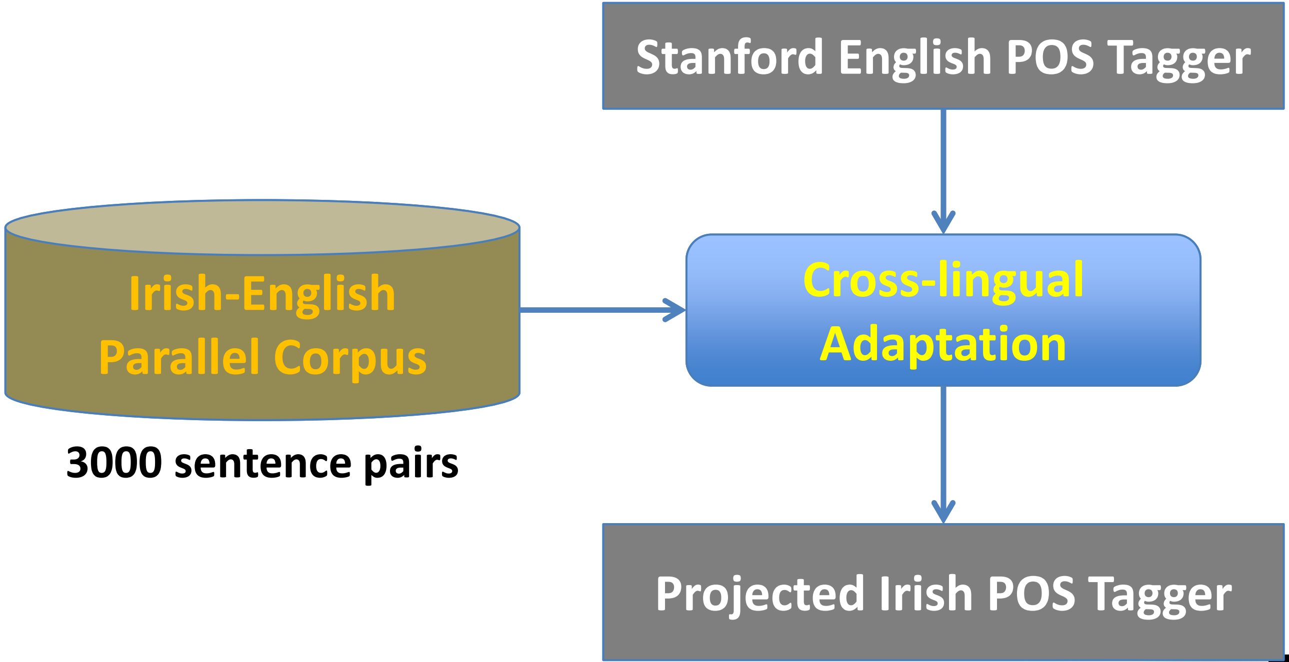
- Irish Dependency Treebank with POS tags: 1022 trees
  - Test set: top 100 trees
  - Development set: next 100 trees
  - Training set: other 822 trees

- Irish-English parallel corpus: 65005 sentence pairs
  - Irish: 1,257,153 tokens
  - English: 1,102,908 tokens

Baseline Irish POS-Tagger

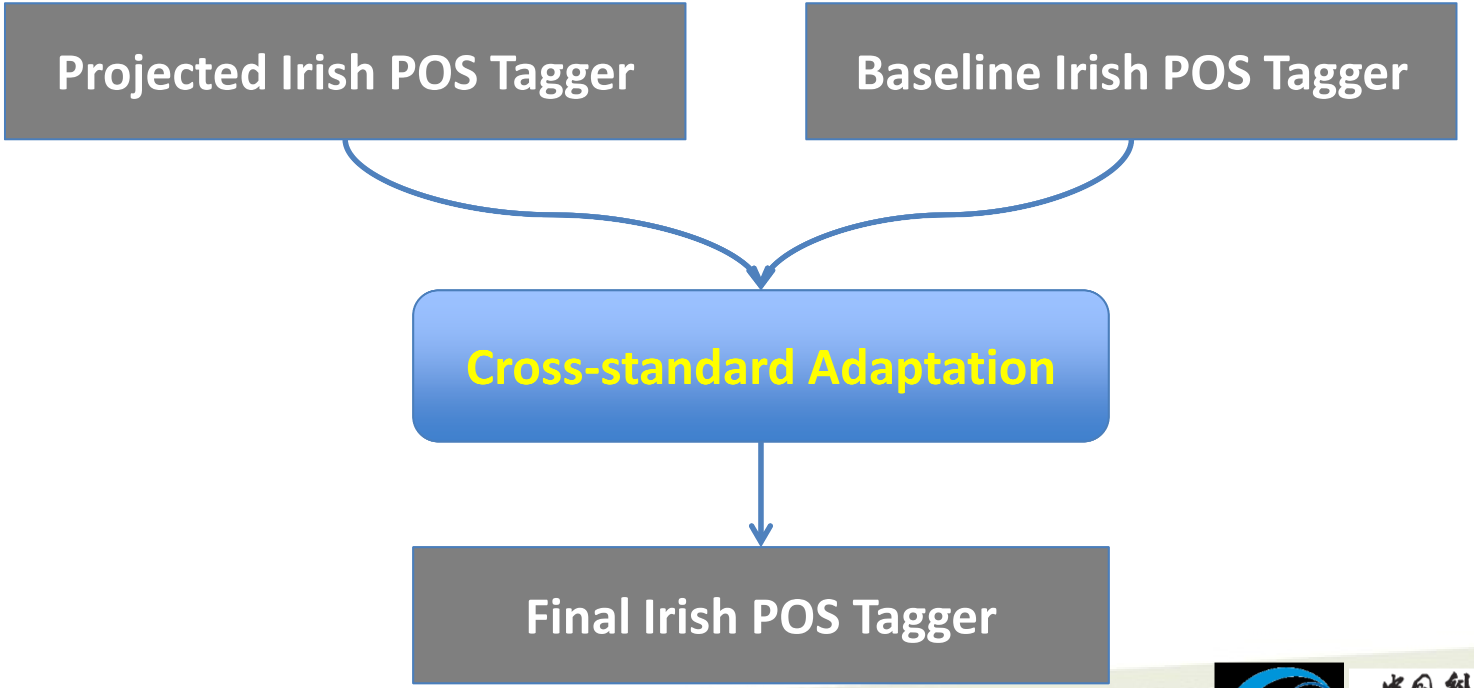


Projected Irish POS Tagger

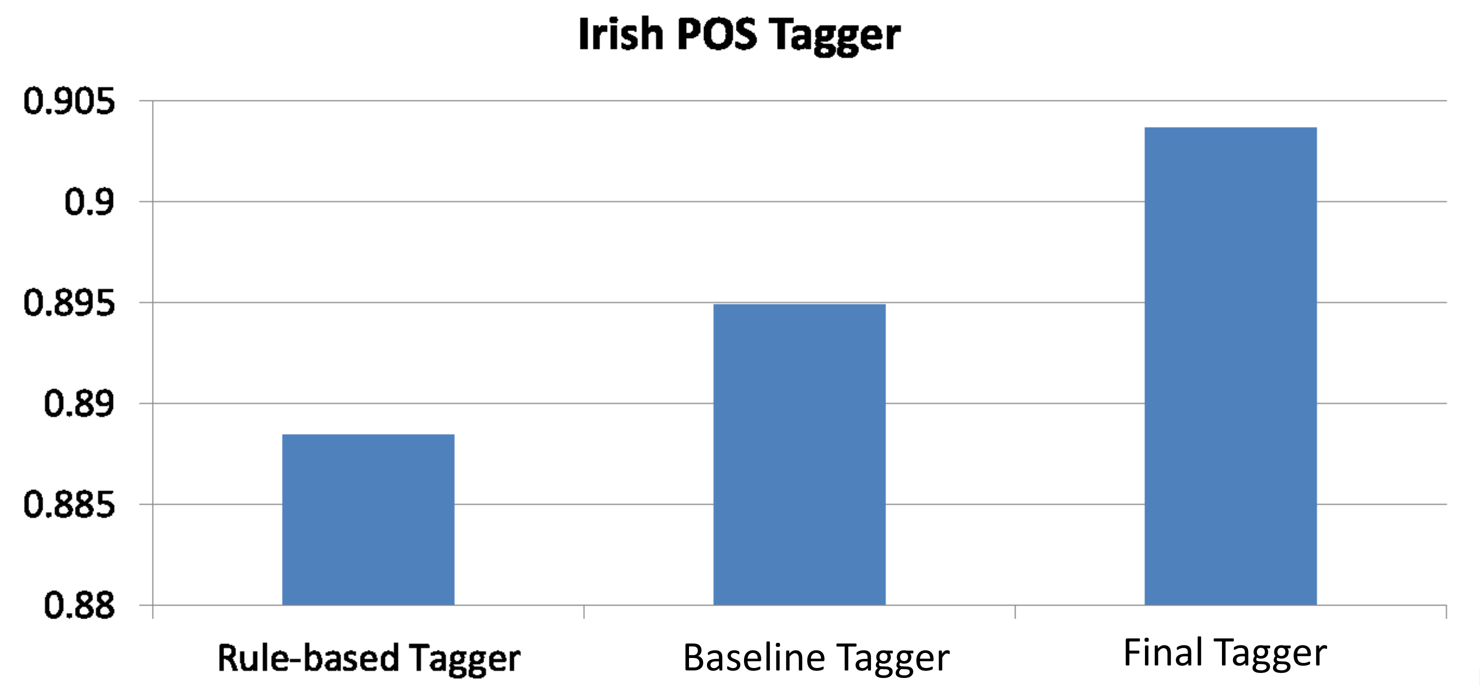


3000 sentence pairs

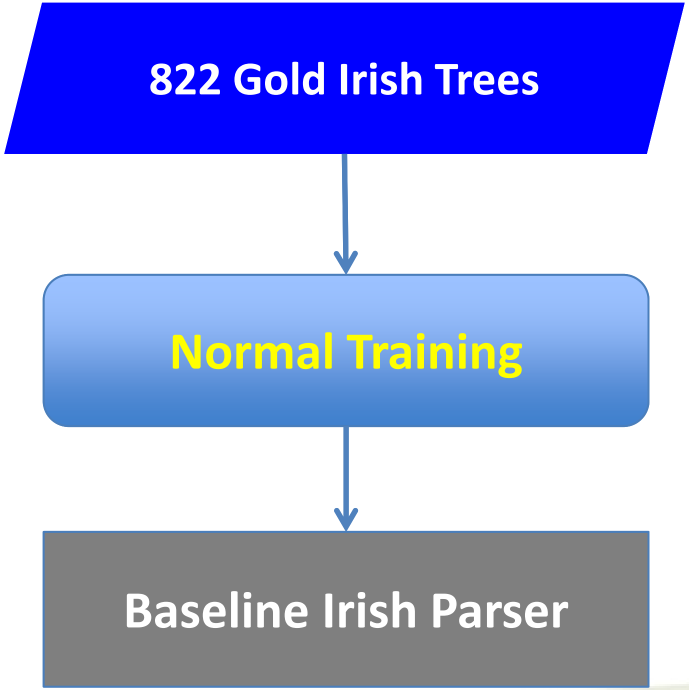
Final Irish Parser



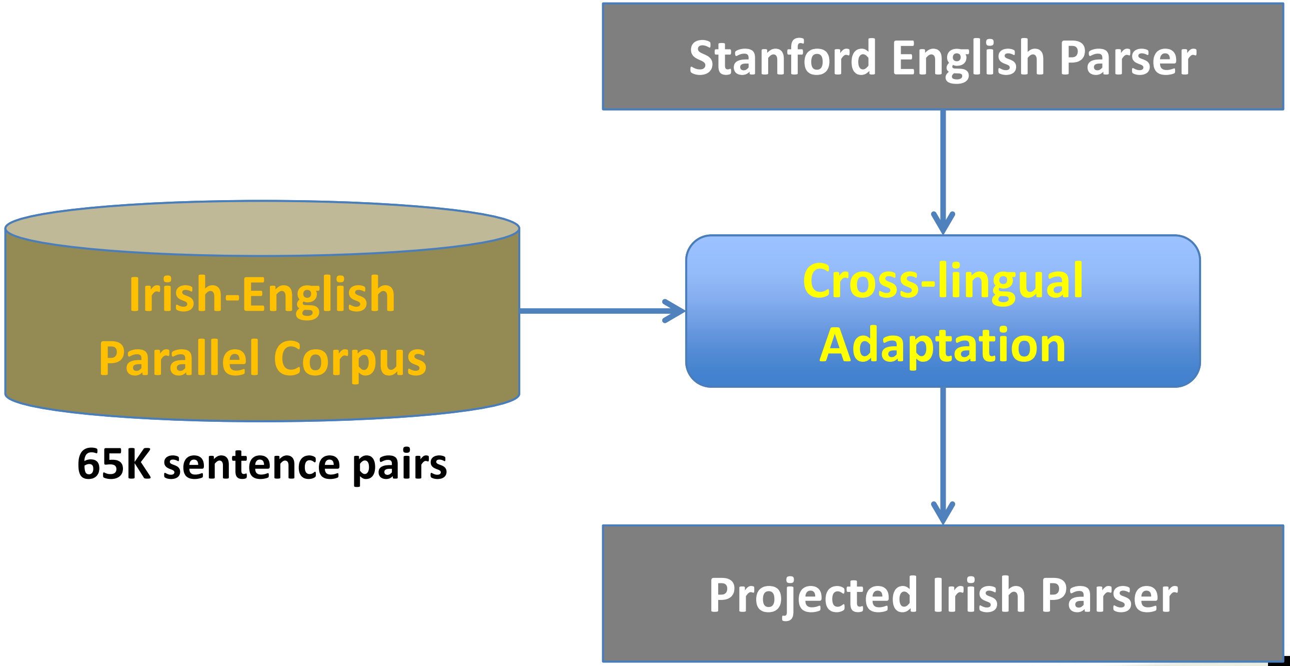
Results of Irish POS Tagging Adaptation



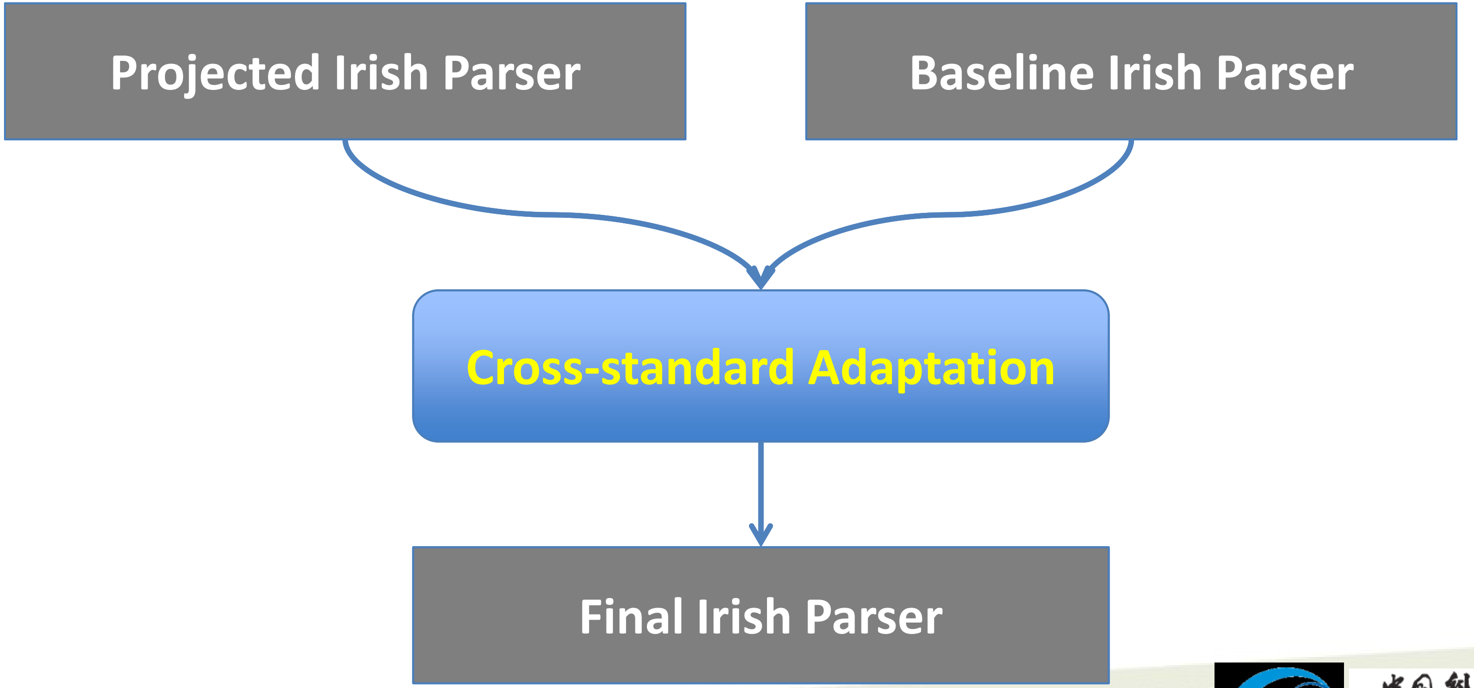
Baseline Irish Parser



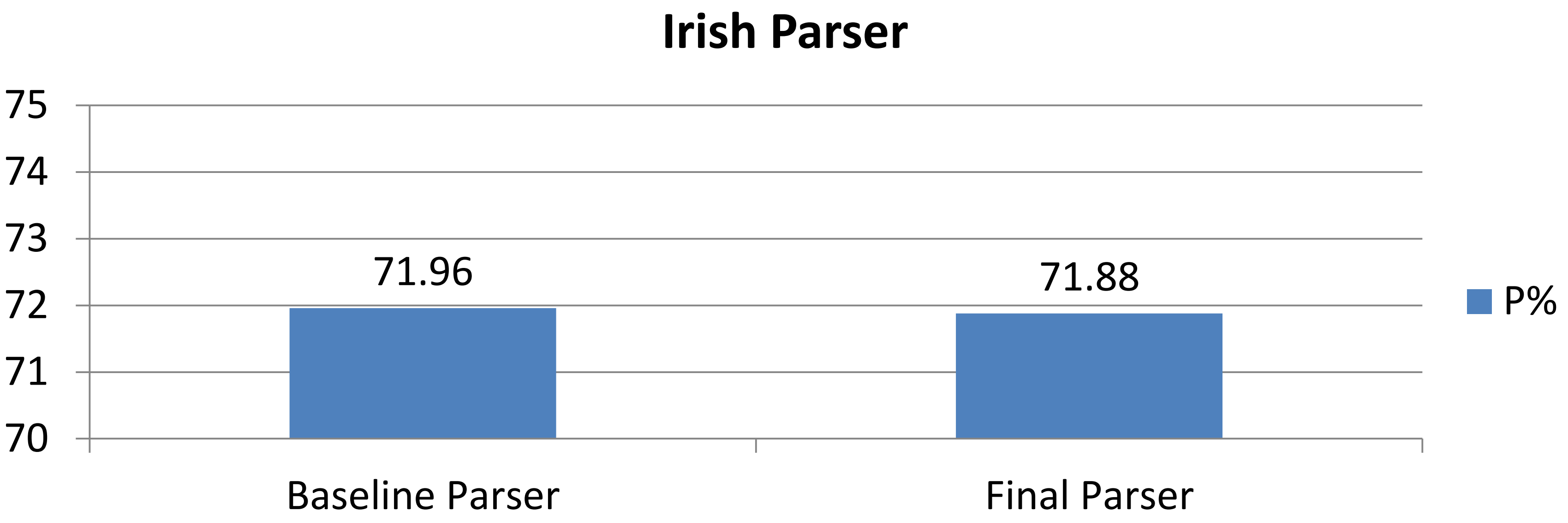
Projected Irish Parser



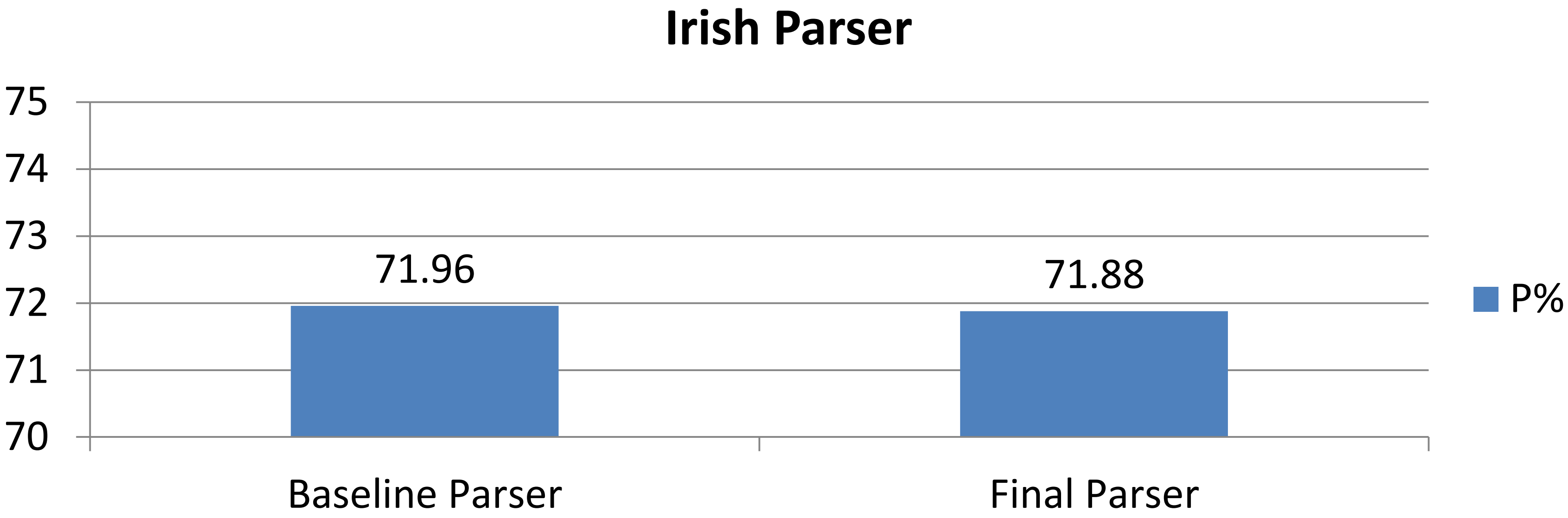
Final Irish Parser



Experiments – Standard Settings

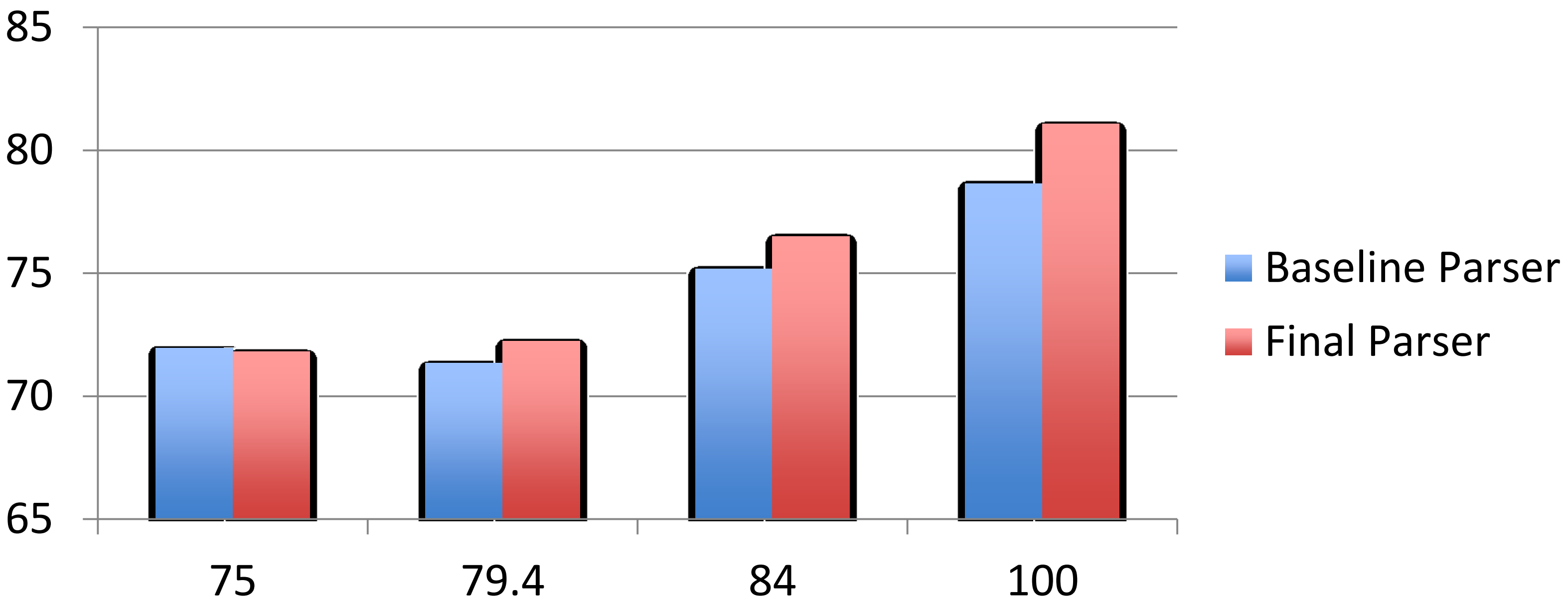


Experiments – Standard Settings



Coverage of word in test set by parallel corpus: 75%

Improvement vs. Test Data Coverage



We re-splitted the training set and test set to make the test set has a higher word coverage by the parallel corpus.

Summary on Irish Experiments

- We conducts joint cross-lingual adaptation and cross-standard adaptation on Irish POS tagging and dependency parsing
- Our results outperform the state-of-the-art Irish POS tagger and parser
- The improvement of adaptation dependents on the coverage of the words in the test set by the bilingual corpus

- Question: is it possible solve the word coverage problem by using domain adaptation technology?

Outline

- Introduction
- Cross-Standard Adaptation
- Cross-Lingual Adaptation
- Experiments on Irish Processing
- **Conclusion**

Conclusion

- Data scarcity is a problem for NLP forever
- Adaptation is a promising technology to alleviate the data scarcity problem
- We proposed two novel technologies:
  - Conditional Mapping for Cross-standard Adaptation
  - Decomposed Projection for Cross-lingual Adaptation
- These two technologies are used to solve the adaptation for Chinese word segmentation and dependency parsing and our results outperform the stat-of-the-art work.
- Latest experiments on Irish POS tagging and dependent parsing also show significant improvements on very strong baselines.

Conclusion

Whenever we have data scarcity problem:

Let’s Adapt!

Acknowledgement

- **SFI** (Ireland) – CNGL II
- **MOST** (China) – "863" projects
- Dr. **Kai Liu**
  - Cross-lingual syntax projection & unsupervised EM training
- Mr. **Jian Zhang**, Ms. **Teresa Lynn**, Dr. **Jennifer Foster**
  - Irish processing
- Dr. **Elaine Uí Dhonnchadha**
  - Irish corpus and rule-based POS tagger
- Dr. **John Moran**, Ms. **Teresa Lynn**, Dr. **John Judge**
  - Irish-English parallel corpus
- Dr. **Jennifer Foster,** Prof. **Vincent Wade,** Mr. **Chris Hokamp,** Prof. **Andy Way,** Mr. **Piyush Arora**
  - Comments and suggestions on slides preparation and presentation

Thank you!