

# Machine Translation Selection

Kaili Zhang

12/18/2013

# Motivation

Google

+Teng



Share



0.1K/S

Translate

English Spanish French Detect language



Chinese (Simplified)

English

Spanish

Translate

Type text or a website address or [translate a document](#).

bing

TRANSLATOR

FOR PHONE

FOR PC

FOR WEBMASTERS

FOR BING BAR

HELP

Translate from: Auto-Detect

Translate to: English



Translate

Baidu 翻译 在线翻译

源语言: 自动检测



目标语言: 英语

翻译

请输入您要翻译的文字内容或网页地址

Translate with SYSTRAN

Text Web Page

French English

Translate

有道 youdao 免费、即时的多语种在线翻译

原文/网址:

# Feature Vector

-- represent each result

1. Translation Engine's Bleu Score

----- Whole Impression, Engine Level

2. TER(translation error rate)

----- Vote Feature, Sentence Level

3. Name Entity

----- Preserved Token, Word Level

4. Dependency Relation

----- Preserved dependency, Word Level

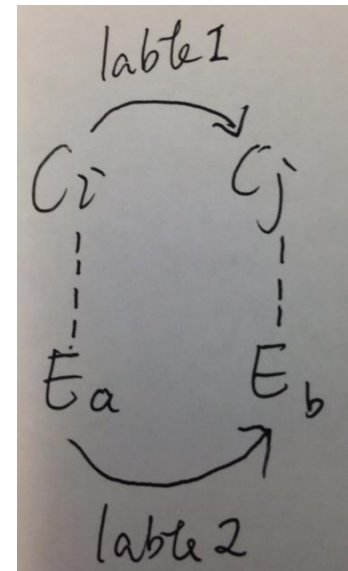
## Dependency Relation

Define:  $C_i$ ,  $C_j$  are two Chinese tokens in Chinese sentence  
 $E_a$ ,  $E_b$  are two English tokens in English sentence

If:  $\text{arch}(C_i \rightarrow C_j) \text{ EXIST}$

if:  $C_i$  is aligned to  $E_a$   
 $C_j$  is aligned to  $E_b$

if:  $\text{arck}(E_a \rightarrow E_b) \text{ EXIST}$



conclusion: this dependency relation preserved.

# Feature Vector

-- represent each result

5. Language Model

----- Sentence Level

6. Word Penalty

----- Sentence Level



Model

Log Linear Algorithm

# Experiment

Traning Data: 422 Sentences from Gale

Testing Data: 422 Sentences from Gale

Tools: Moses  
Stanford Dependency Parser  
Stanford Named Entity  
Stanford Tagging

OFFLINE

Engines' Best = 32.49

Feature Combination	BLEU on Testing Set
LM , WP	29.72
Dependency + LM, WP	30.53
Labeled Dependency + LM, WP	29.81
Named-Entity + LM, WP	30.86
Dependency + Named-entity + LM, WP	30.89
TER + LM, WP	32.17
BLUE + LM, WP	32.49
ALL	32.52



ENGINE	BLEU on Testing Set
Google	25.98
Bing	19.76
Youdao	14.40
Selection Model (BLEU + TER + LM + WP)	26.09