Statistical Syntax-Directed Translation

with Extended Domain of Locality

Liang Huang (Penn), Kevin Knight (ISI), Aravind Joshi (Penn)

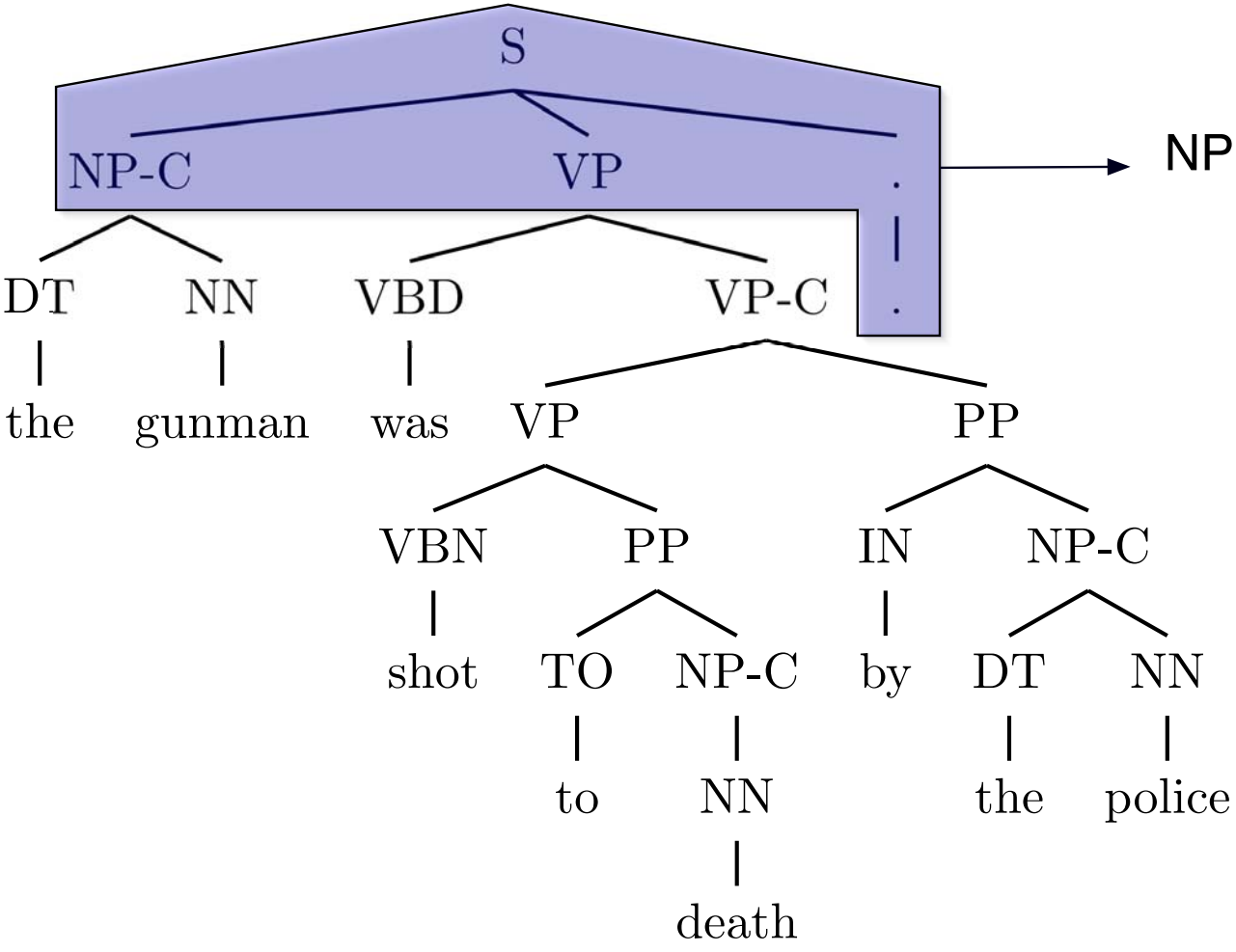
AMTA 2006, Boston, MA

Syntax-Directed Translation

- How do we human-beings do translation?
  1. understand the source sentence
  2. generate the target sentence
- Compiling
  1. parse input program into an expression/syntax tree
  2. generate code in machine language
- Machine Translation?
  1. parse the source sentence into a tree
  2. recursively transfer the tree into the target language
     - *this work*

Recursive Transfer

- converting tree-fragments into the target language

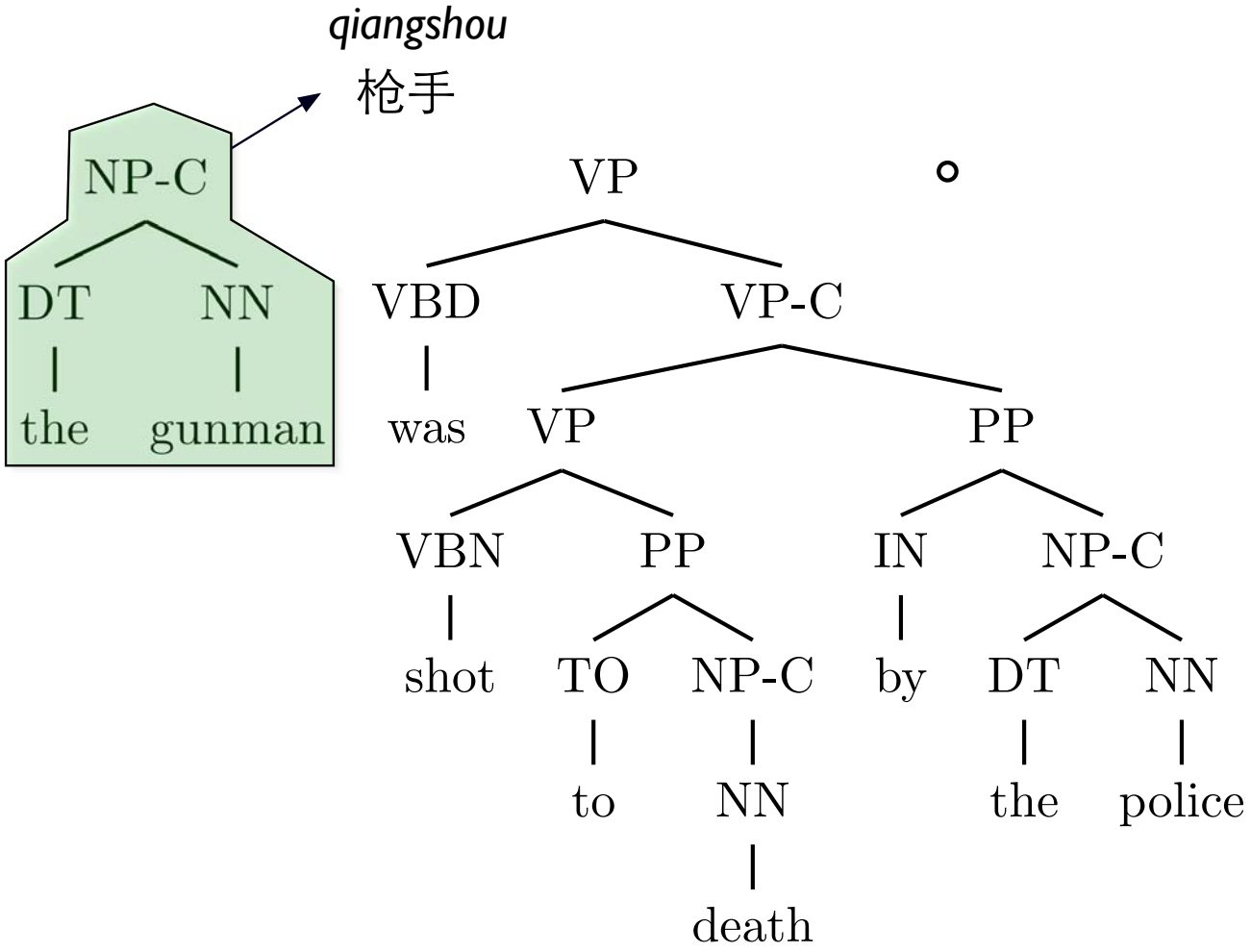


NP-C VP 。

using the rules of
(Galley et al., 2004)

Recursive Transfer

- converting tree-fragments into the target language



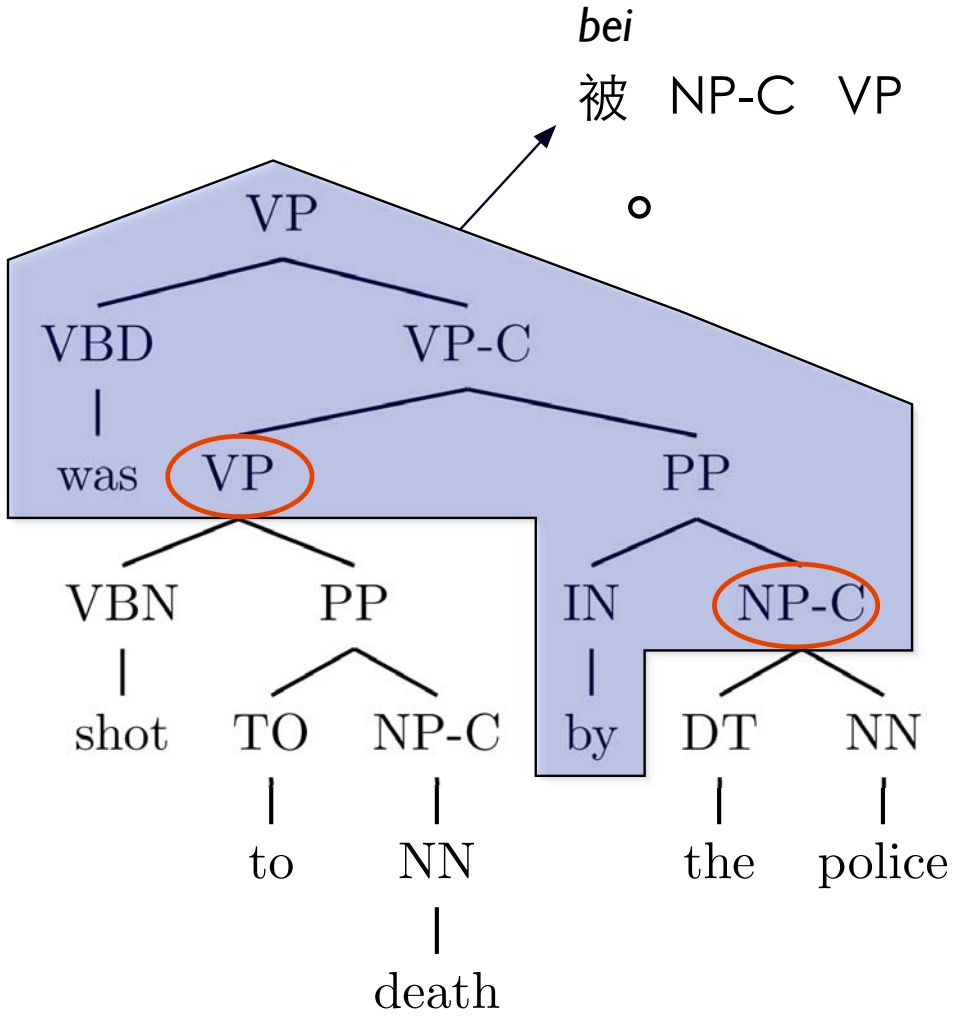
Recursive Transfer

- converting tree-fragments into the target language

枪手

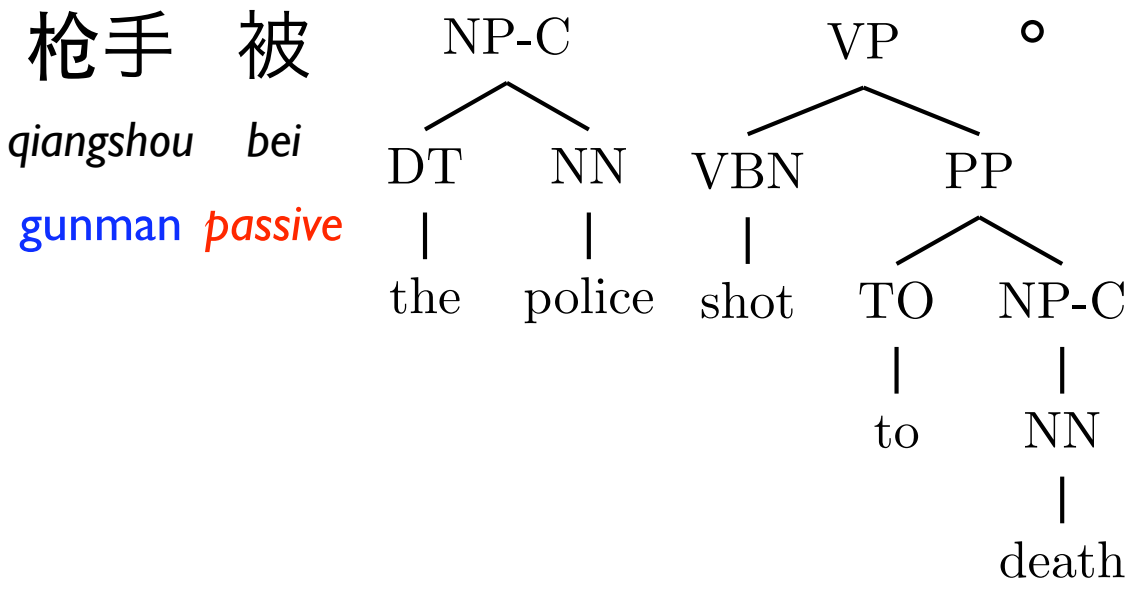
qiangshou

gunman



Recursive Transfer

- converting tree-fragments into the target language



Recursive Transfer

- converting tree-fragments into the target language

枪手	被	警察	击毙	。
qiangshou	*bei*	*jingcha*	*jibi*	
gunman	*passive*	police	kill	

formal framework (Graehl and Knight, 2004)

- 1-state
- extended left-hand-side
- linear and non-deleting
- root-to-frontier tree transducer
- (1-xRLN)

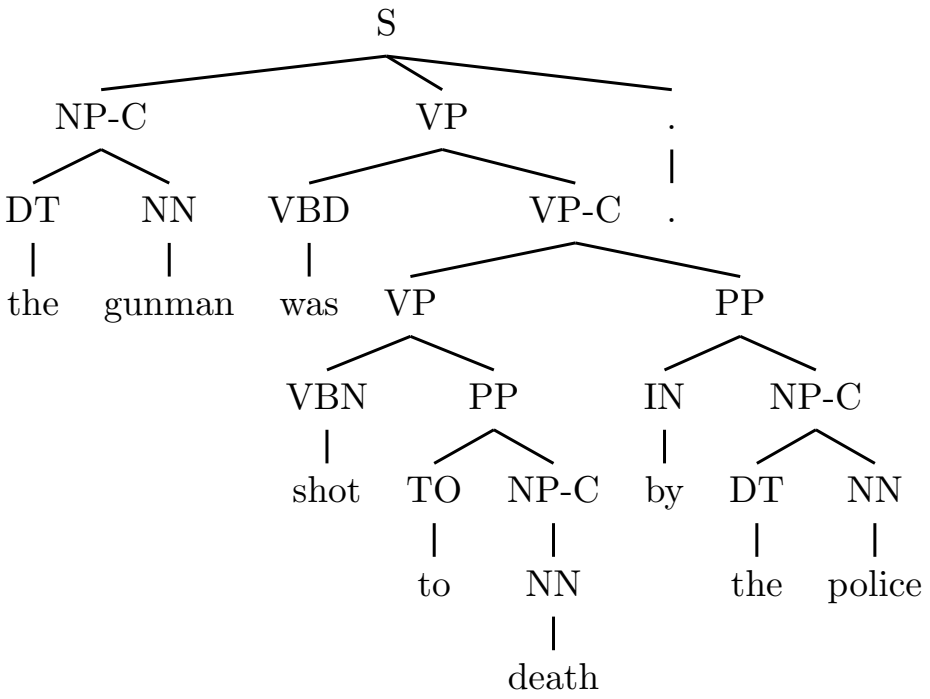
Comparison with ISI system

枪手被警方击毙。

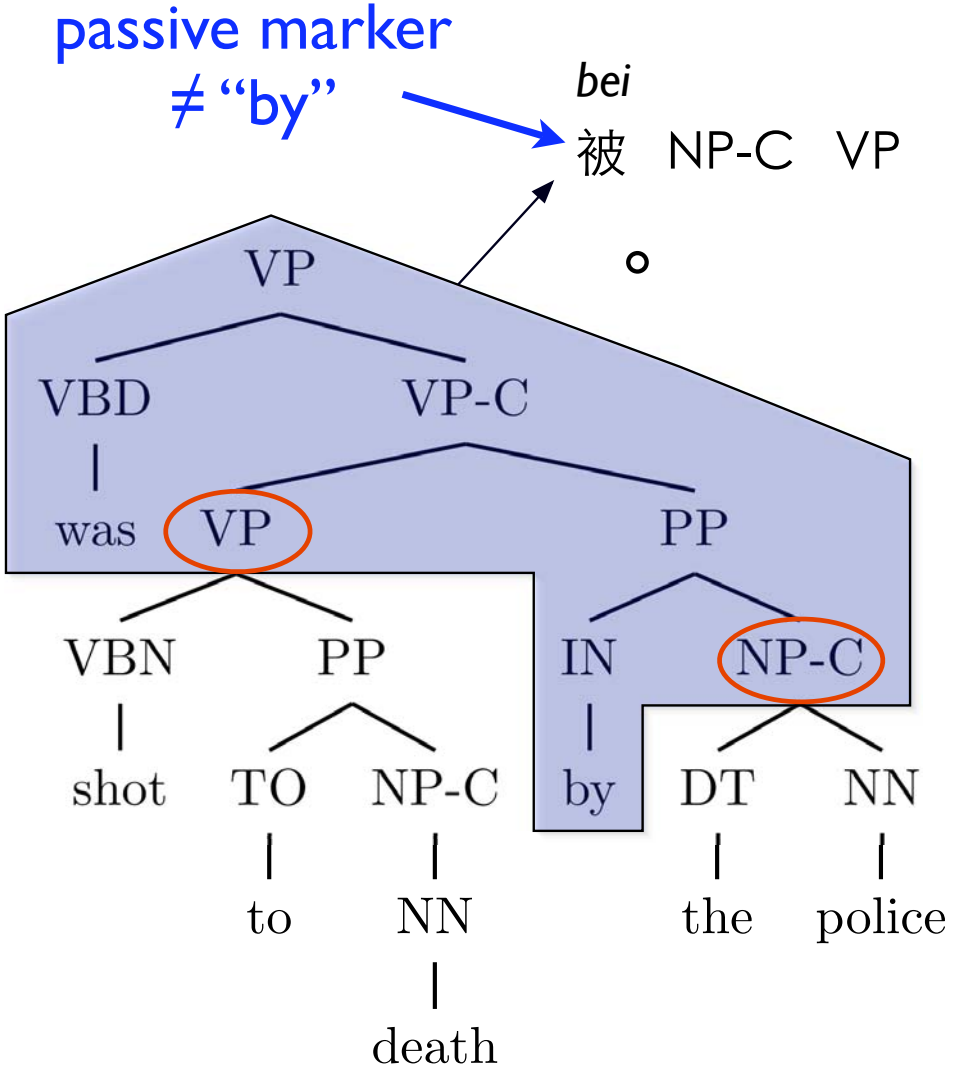
ISI: decoding

ours: encoding

The gunman was killed by police.



Comparison with SCFGs



(Chiang, 2005):

(was X_1 by X_2, 被 X_2 X_1)

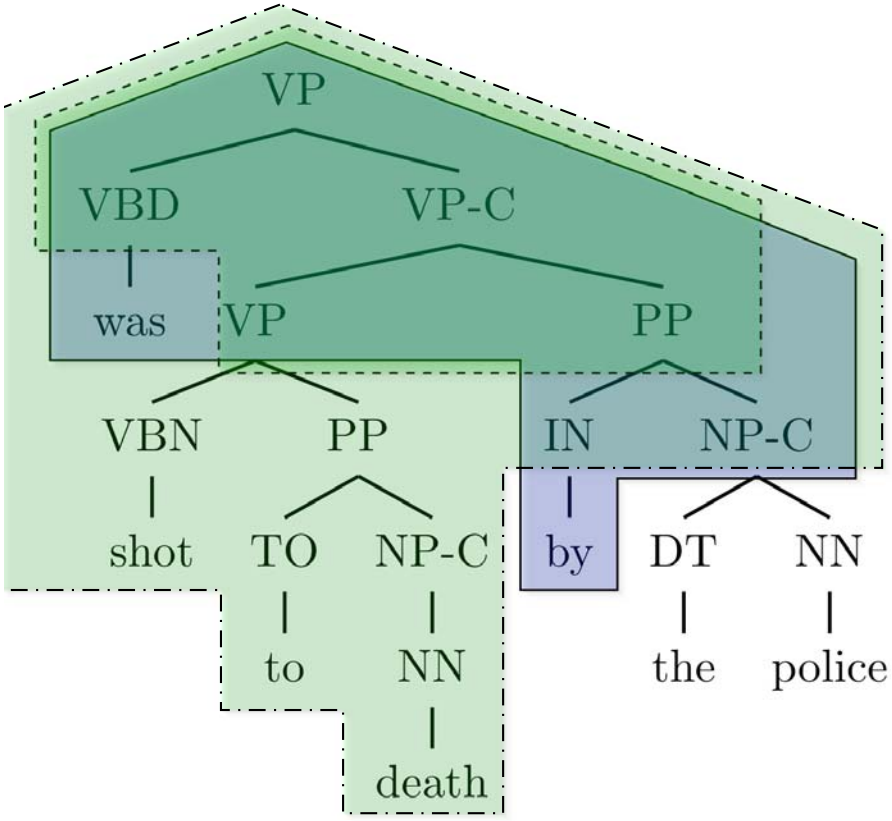
I was here by myself

我 被 我 自己 在 这里

I was asleep by 10 pm

我 被 10 点 睡着

Search



- depth-first-search (DFS)
  - for each tree node
  - try all rules applicable
  - recursion on subtrees
  - plugin the results
- many overlapping rules
  - exponential run-time!
- solution: memoization
  - every node visited once
  - linear-time algorithm

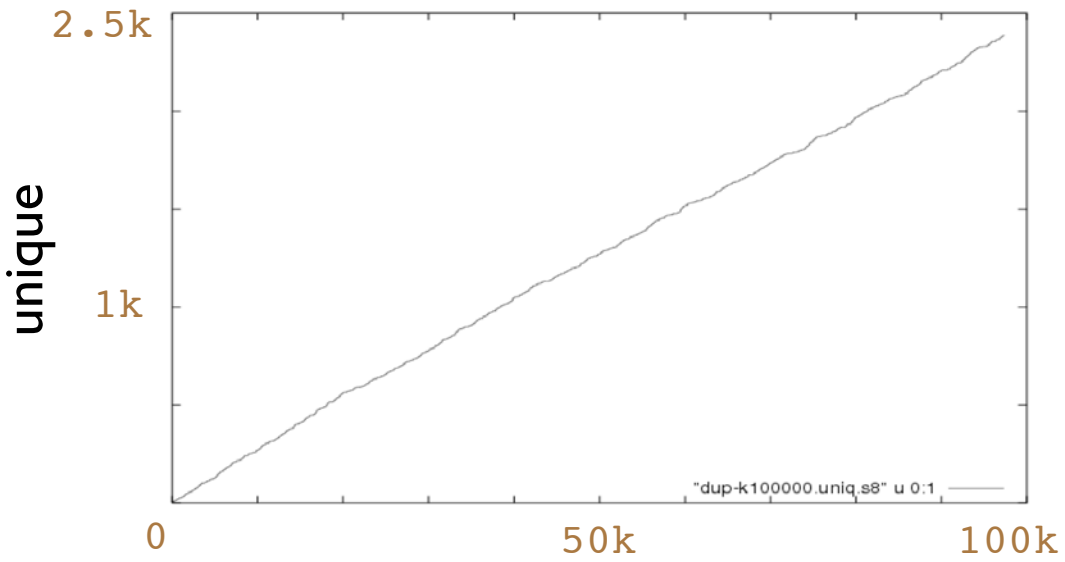
Adding Language Models

- Generate top-*k* translations
  - then rescore with language model
- fast, using Algorithm 3 of (Huang and Chiang, 2005)
- but many duplicate translations
  - due to spurious ambiguity
  - one solution: determinizion (May and Knight, 2006)

1 110.901 目前 , 一 些 西方 国家 已 公布 终止 津巴布韦 经济援助 。
2 111.042 目前 , 一 些 西方 国家 已 宣布 终止 津巴布韦 经济援助 。
3 111.101 目前 , 一 些 西方 国家 已 公布 终止 津巴布韦 经济援助 。
4 111.142 目前 , 一 些 西方 国家 已 宣布 终止 津巴布韦 经济援助 。
5 111.264 目前 , 一 些 西方 国家 已经 公布 终止 津巴布韦 经济援助 。
6 111.264 目前 , 一 些 西方 国家 已经 公布 终止 津巴布韦 经济援助 。
7 111.327 目前 , 一 些 西方 国家 已 公布 终止 津巴布韦 经济援助 。
8 111.361 目前 , 一 些 西方 国家 已经 公布 终止 津巴布韦 经济援助 。

Non-Duplicate *k*-best

- a simple modification to the lazy k-best algorithm
  - at each node, store a list of "unique solutions"
  - keep asking for next-best until you get something new
  - duplicates eliminated recursively at sub-problems
  - here "uniqueness" can be any equivalence relation



duplicate ratio = 1:40

with our trick:

getting 100k unqiue is only about twice as slow as 100k non-unique

all

Experiments: English-to-Chinese

- Training data
  - Chinese/English parallel data from newswire
  - 1.95 M sentence pairs, English side: 28.3 M tokens
  - 24.7 M tree-to-string rules extracted (Galley et al., 2004)
  - Chinese trigram model trained on the Chinese side
- Evaluation data
  - NIST 2003 evaluation set (Chinese-to-English)
  - subset: 140 short sentences (≤ 25 Chinese words)
  - pick the first and second English references as source
  - equally divide into dev and test (each w/ 140 sent.)
  - single Chinese reference for each input sentence

Systems

- Baseline: Pharaoh (Koehn, 2004)
  - log-linear model of eight features
  - feature weights tuned by max-bleu algorithm (Och, 2003)
- This work
  - build the TM derivation forest (linear-time algorithm)
  - get top-*k* unique translations
  - rescore with the trigram model
  - log-linear model of 3 features: TM, LM, length penalty
  - fix w_{TM} , grid-search w_{LM} , and binary-search len-penalty

Results

- character-based BLEU (Chinese output), 95% interval

System	dev set BLEU-4	test set (140 sentences) BLEU-4	BLEU-8
Pharaoh (with max-BLEU tuning)	25.96 ±2.8	23.54 ±1.9	6.739 ±1.2
direct model (1-best)	22.10 ±2.6	24.53 ±2.2	7.309 ±1.9
log-linear model (rescoring non-duplicate k-best list)			
$k = 5000$ $(\beta = 0.994, \lambda = 0.513)$	26.01 ±2.7	25.74 ±2.3	8.489 ±2.1
$k = 50000$ $(\beta = 0.793, \lambda = 0.469)$	26.95 ±2.8	26.69 ±2.4	9.323 ±2.2

p<0.05 p<0.01

sign-test (Collins et al., 2005) over Pharaoh

both BLEU-4 and BLEU-8 significantly better than Pharaoh on test-set

Examples

source	the small town of wertheim is expected to experience the highest water level in 80 years .
reference	小镇 韦 尔 特 海 姆 预计 将 经历 8 0 年 来 的 最高 水位 。
pharaoh	小 城镇 , 预计 最高 水位 。 80 年 的 经验 韦 特 海 姆
1-best TM	韦 尔 特 海 姆 小镇 预计 经历 最高 水位 在 80 年 。
rescored	韦 尔 特 海 姆 小镇 将 经历 80 年 的 最高 水位 。

source	the global economic rebound and sustained economic growth in china last year have infused fresh vitality into the development of the port of shanghai .
reference	去年 全球 经济 回 暖 , 中国 经济 持续 发展 , 给 上海 港 的 发展 注入 新 的 活力 。
pharaoh	世界 经济 的 回升 , 经济 持续 增长 注入 新 的 活力 中国 去年 的 港口 发展 到 上海 。
1-best TM	去年 中国 全球 经济 回升 与 经济 持续 增长 在 有 注入 新 的 活力 为 上海 港口 的 发展 。
rescored	去年 中国 全球 经济 回升 和 经济 持续 增长 为 上海 港口 的 发展 注入 新 的 活力 。