

Statistical NLP Spring 2008

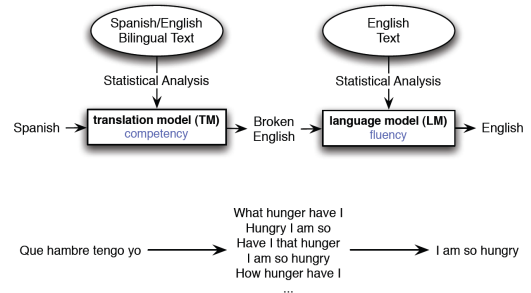


Lecture 25: Syntactic Translation

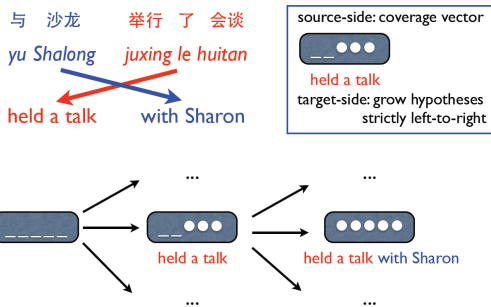
Dan Klein – UC Berkeley

Slides from Liang Huang and Jonathan May

MT Overview

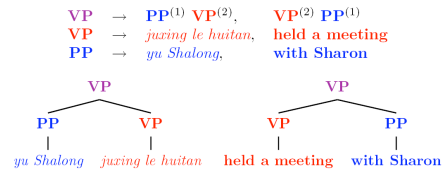


Phrase-Based MT



Syntax-Based MT

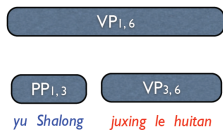
- synchronous context-free grammars (SCFGs)
- context-free grammar in two dimensions
- generating pairs of strings/trees simultaneously
- co-indexed nonterminal further rewritten as a unit



Translation by Parsing

- translation with SCFGs => monolingual parsing
- parse the source input with the source projection
- build the corresponding target sub-strings in parallel

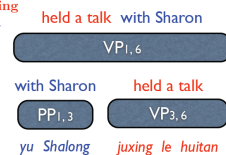
$VP \rightarrow PP^{(1)} VP^{(2)}$
 $VP \rightarrow juxing\ le\ huitan,$
 $PP \rightarrow yu\ Shalong,$

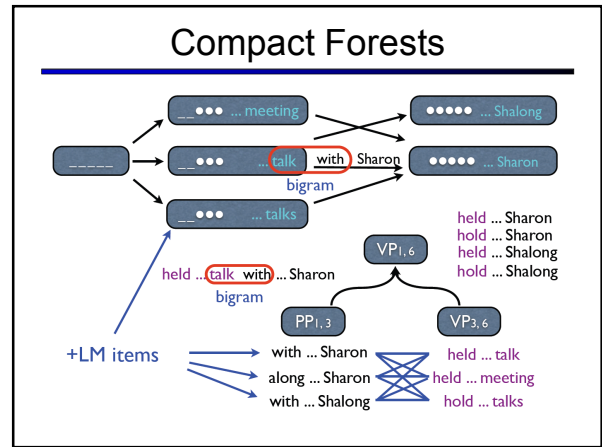
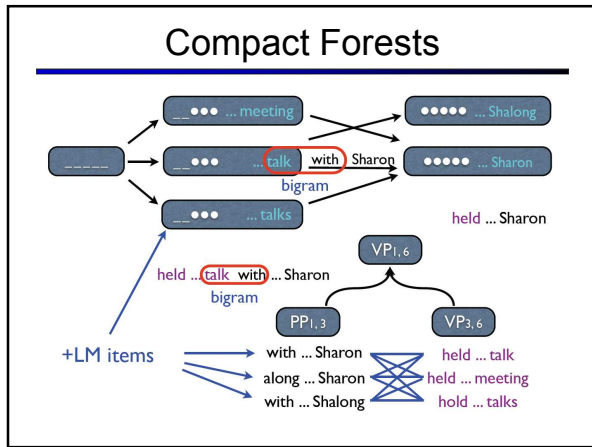
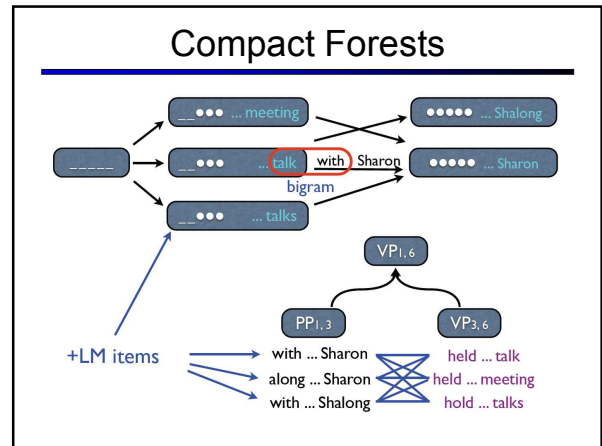
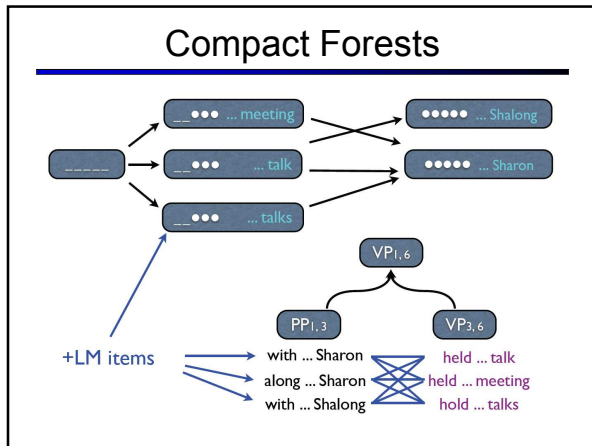


Translation by Parsing

- translation with SCFGs => monolingual parsing
- parse the source input with the source projection
- build the corresponding target sub-strings in parallel

$VP \rightarrow PP^{(1)} VP^{(2)}$, $VP^{(2)} PP^{(1)}$
 $VP \rightarrow juxing\ le\ huitan,$ $held\ a\ meeting$
 $PP \rightarrow yu\ Shalong,$ $with\ Sharon$





Language Model Costs

non-monotonic grid due to LM combo costs

	1.0	3.0	8.0	
(VP _{3,6} ^{held} * meeting)	1.0	2.0 + 0.5	4.0 + 5.0	9.0 + 0.5
(VP _{3,6} ^{held} * talk)	1.1	2.1 + 0.3	4.1 + 5.4	9.1 + 0.3
(VP _{3,6} ^{held} * conference)	3.5	4.5 + 0.6	6.5 + 10.5	11.5 + 0.6

Language Model Costs

k-best parsing (Huang and Chiang, 2005)

- a priority queue of candidates
- extract the best candidate
- push the two successors

	1.0	3.0	8.0	
(VP _{3,6} ^{held} * meeting)	1.0	2.5	9.0	9.5
(VP _{3,6} ^{held} * talk)	1.1	2.4	9.5	9.4
(VP _{3,6} ^{held} * conference)	3.5	5.1	17.0	12.1

Language Model Costs

[Huang and Chiang 06]

items are popped out-of-order
solution: keep a buffer of pop-ups

2.5 2.4 5.1

finally re-sort the buffer
 and return in order:

2.4 2.5 5.1

(VP held * meeting)
 3,6

(VP held * talk)
 3,6

(VP hold * conference)
 3,6

	1.0	3.0	8.0
1.0	2.5	9.0	9.5
1.1	2.4	9.5	9.4
3.5	5.1	17.0	12.1

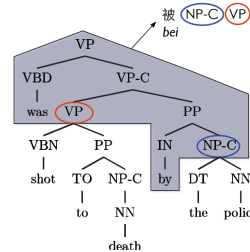
(PP with * Sharon)
 1,3

(PP along * Sharon)
 1,3

(PP with * Sharon's)
 1,3

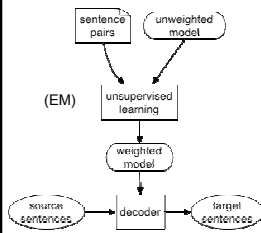
Learning MT Grammars

- syntax-directed, English to Chinese (Huang, Knight, Joshi, 2006)
- first parse input, and then recursively transfer



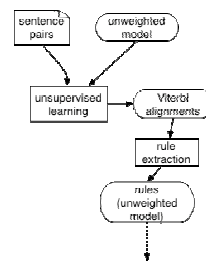
synchronous tree-substitution grammars (STSG)
 (Galley et al., 2004; Eisner, 2003)

Idealistic machine translation



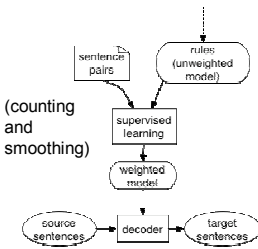
- Theorize a translation model
- Train on parallel sentences with unsupervised learning methods
- Examples: Berger et. al '94 (IBM Model 4), Wu '96 (ITG), Yamada and Knight '01

Realistic machine translation



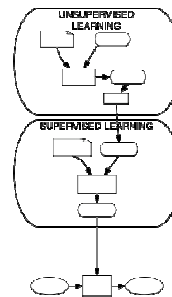
- A relatively simple model is used to learn alignments unsupervisedly
- The alignments are used to limit the exploration space of a more complicated model

Realistic machine translation



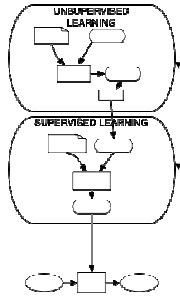
- The complicated model is trained on a smaller set of sentence pairs
- Counting and smoothing is fairly quick

Realistic machine translation



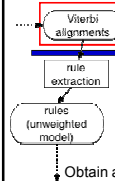
- Empirically shown to be better than idealistic approach
- Scales well
- Examples: Och & Ney '04, Chiang '05, Galley et. al '06

Realistic machine translation

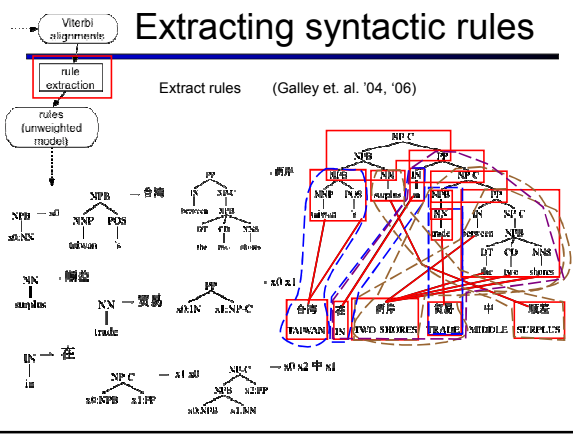


- New research is on advanced models, but older models are used as bootstrap
- Can we add the power in our supervised model to the unsupervised component?

Extracting syntactic rules

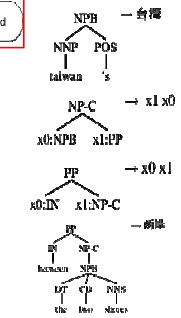
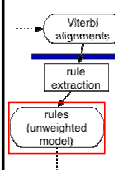


Extracting syntactic rules



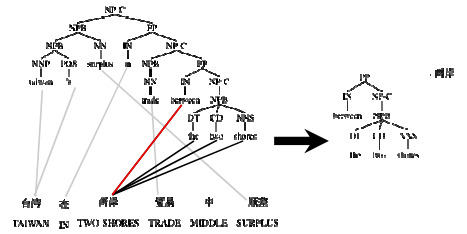
Extract rules (Galley et. al. '04, '06)

Rules can...



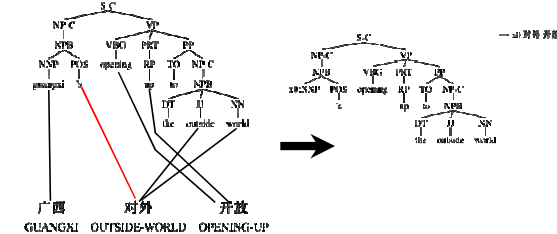
- capture phrasal translation
- reorder parts of the tree
- traverse the tree without reordering
- insert (and delete) words

Bad alignments make bad rules



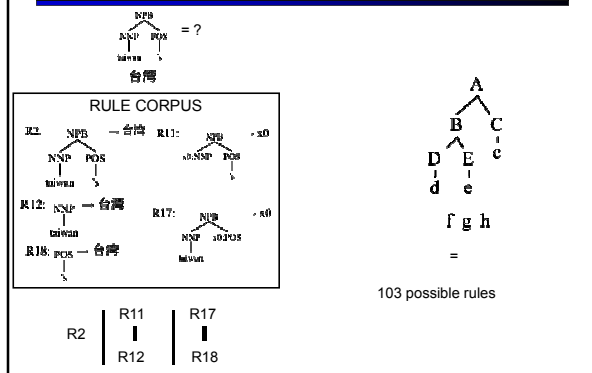
This isn't very good, but let's look at a worse example...

Sometimes they're really bad

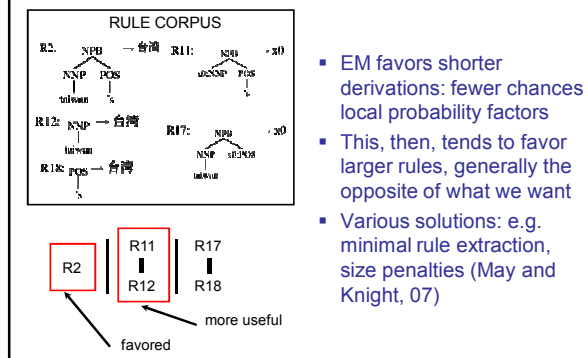


One bad link makes a totally unusable rule!

Rule Learning?



EM size bias



Other Important Issues

- Rule composition
- Tree transforms (binarization, etc)
- Crazy pruning

Results

From [DeNeefe et al 07]

Experiment	Chinese		Arabic	
	Dev	Test	Dev	Test
Baseline ATS	34.94	32.83	50.46	50.52
Baseline GHKM (minimal only)	38.02	37.67	49.34	49.99
GHKM composed size 2	40.24	39.75	50.76	50.94
GHKM composed size 3	40.95	40.44	51.56	51.48
GHKM composed size 4	41.36	40.69	51.60	51.71
GHKM minimal + SPMT model 1	39.78	39.16	50.17	51.27
GHKM composed + SPMT model 1	42.04	41.07	51.73	51.53
With binarization	42.17	41.26	52.50	51.79