CS 294-5: Statistical Natural Language Processing

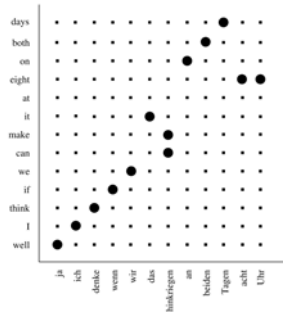
Machine Translation II

Dan Klein

includes slides from Yamada, Knight, Koehn

Assignment 2 Honors

“Shape” of Alignments



IBM Models 1+2

- Japan shaken by two new quakes

Le Japon secoué par deux nouveaux séismes

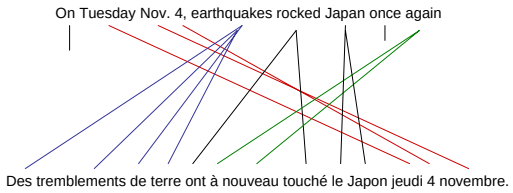
Learning with EM

- Model 1 Parameters: P(f|e)
- Start with P(f|e) uniform, including P(f|null)
- For each sentence:
  - For each French position i
    - Calculate posterior over English positions, $P(a_i|i)$

$$P(a_i \mid i) = \frac{P(f_i \mid e_{a_i})}{\sum_{a_{i'}} P(f_i \mid e_{a_{i'}})}$$

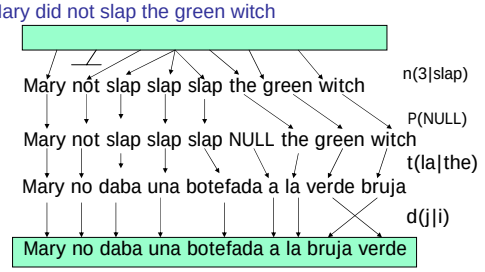
    - Increment count of word f_i with word e_{ai}
- Iterate until convergence

HMM Alignment Model



IBM Model 3

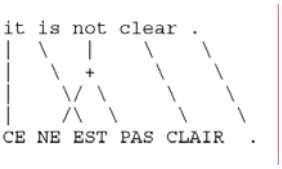
- Mary did not slap the green witch



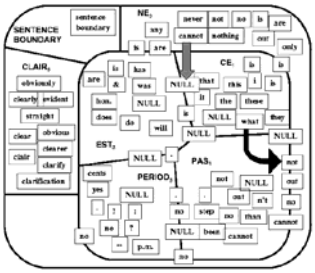
[Al-Onaizan and Knight, 1998]

Decoding

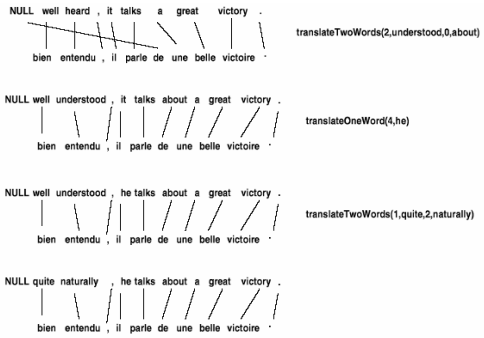
- In these word-to-word models
  - Finding best alignments is easy
  - Finding translations is hard (why?)



Exact Decoding as a TSP



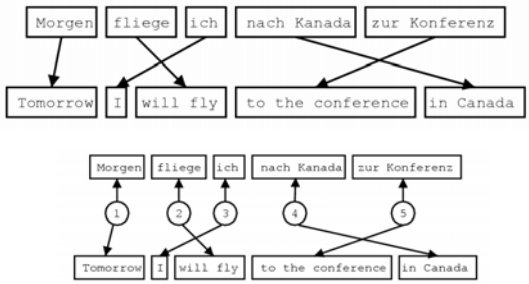
Greedy Decoding



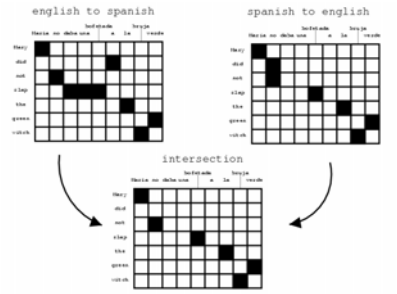
Stack Decoding

sent length	decoder type	time (sec/sent)	search errors	translation errors (semantic and/or syntactic)	NE	PME	DSE	FSE	HSE	CE
6	IP	47.50	0	57	44	57	0	0	0	0
6	stack	0.79	5	58	43	53	1	0	0	4
6	greedy	0.07	18	60	38	45	5	2	1	10
8	IP	499.00	0	76	27	74	0	0	0	0
8	stack	5.67	20	75	24	57	1	2	2	15
8	greedy	2.66	43	75	20	38	4	5	1	33

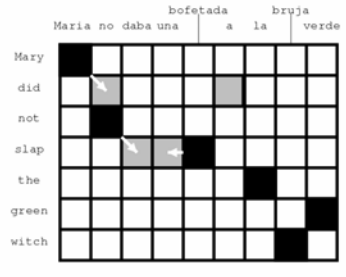
Phrase-Based Systems



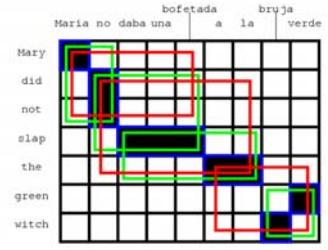
Bidirectional Alignment



Growing Alignments



Extracting Phrases



Models?