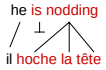
CS 294-5: Statistical Natural Language Processing

Phrase-Based Translation

Lecture 12: 10/17/05

Uses slides from Koehn, Knight et al.

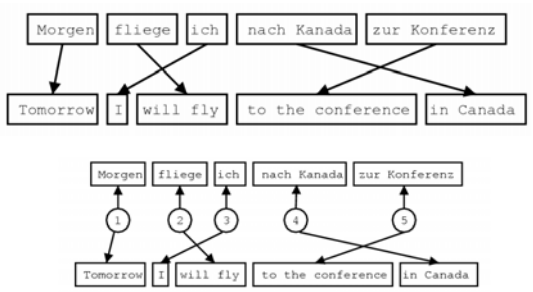
Phrases in IBM Models



nodding

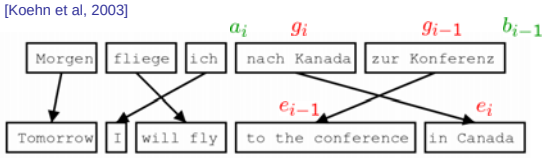
f	$t(f \mid e)$	ϕ	$n(\phi \mid e)$
signe	0.164	4	0.342
la	0.123	3	0.293
tête	0.097	2	0.167
oui	0.086	1	0.163
fait	0.073	0	0.023
que	0.073		
hoche	0.054		
hocher	0.048		
faire	0.030		
me	0.024		
approuve	0.019		
qui	0.019		
un	0.012		
faites	0.011		

Phrase-Based Systems



Pharaoh's Model

[Koehn et al, 2003]



$$P(e|g) = P(\{\bar{g}_i\}|g) \prod_i \phi(\bar{e}_i|\bar{g}_i) d(a_i - b_{i-1})$$

Segmentation Translation Distortion

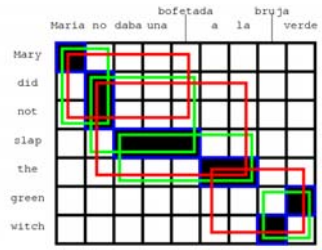
Pharaoh's Model

$$P(f|e) = P(\{\bar{e}_i\}|e) \prod_i \phi(\bar{f}_i|\bar{e}_i) d(a_i - b_{i-1})$$

$\frac{1}{K}$ $\frac{count(\bar{f}_i, \bar{e}_i)}{count(\bar{e}_i)}$ $\alpha^{|a_i - b_{i-1}|}$

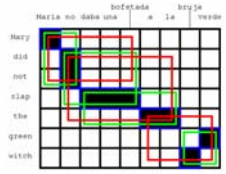
Where do we get these counts?

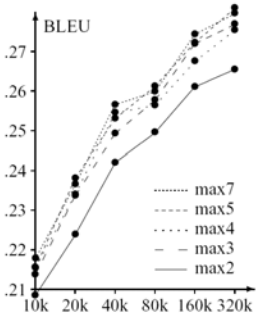
Extracting Phrases



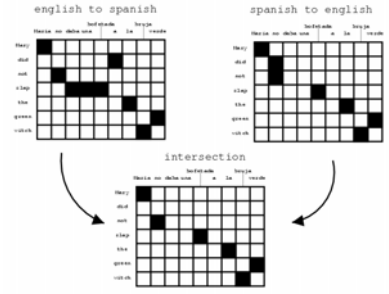
Phrase Size

- Phrases do help
  - But they don't need to be long
  - Why should this be?

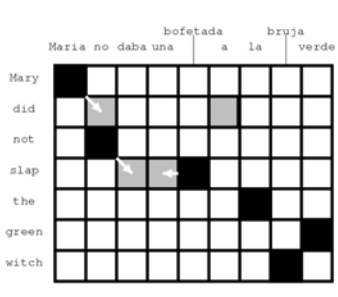


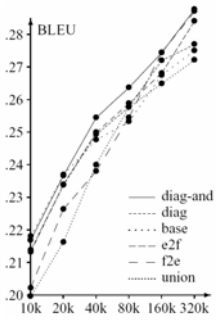


Bidirectional Alignment



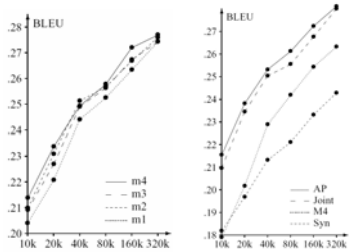
Alignment Heuristics





Sources of Alignments

Method	Training corpus size					
	10k	20k	40k	80k	160k	320k
AP	84k	176k	370k	736k	1536k	3152k
Joint	125k	220k	400k	707k	1254k	2214k
Syn	19k	24k	67k	105k	217k	373k

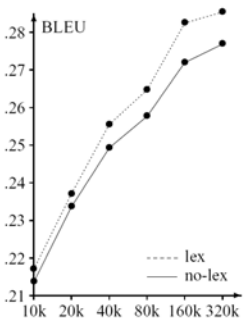


Lexical Weighting

$$\phi(\bar{f}_i|\bar{e}_i) = \frac{count(\bar{f}_i, \bar{e}_i)}{count(\bar{e}_i)} p_w(\bar{f}_i|\bar{e}_i)$$

```
      f1 f2 f3
NULL  -- -- ##
e1    ## -- --
e2    -- ## --
e3    -- ## --
```

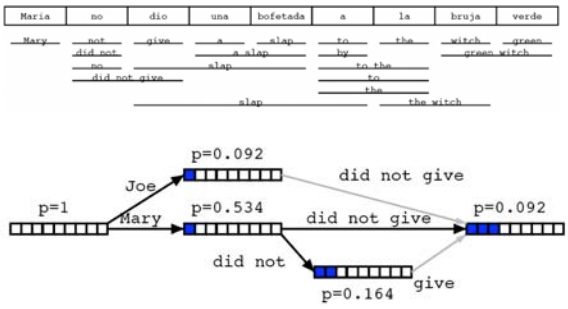
$$\begin{aligned} p_w(\bar{f}|\bar{e}, a) &= p_w(f_1 f_2 f_3|e_1 e_2 e_3, a) \\ &= w(f_1|e_1) \\ &\quad \times \frac{1}{2}(w(f_2|e_2) + w(f_2|e_3)) \\ &\quad \times w(f_3|\text{NULL}) \end{aligned}$$



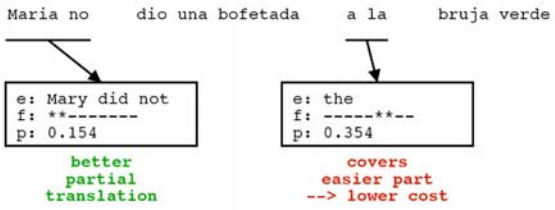
The Pharaoh Decoder

- Probabilities at each step include LM and TM

Hypotheis Lattices



Pruning



- Problem: easy partial analyses are cheaper
  - Solution 1: use beams per foreign subset
  - Solution 2: estimate forward costs (A*-like)

What's Next?

- Modeling syntax
  - PCFGs and phrase structure
  - Syntactic parsing
  - Grammar induction
  - Syntactic language and translation models
- Speech systems
  - Acoustics
  - Applications