Phrase-Based Systems

Pharaoh’s Model

Phrases in IBM Models

Extracting Phrases

Where do we get these counts?
Phrase Size

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

Bidirectional Alignment

Alignment Heuristics

Lexical Weighting

$\phi(f_i)e_i = \frac{\text{count}(f_i,e_i)}{\text{count}(e_i)} \cdot p_n(f_i|e_i)$

$\phi(e_1)$
$\phi(e_2)$
$\phi(e_3)$
$\phi$ (NULL)

Lexical Weighting

$\phi(f_i|e_i) = \frac{\text{count}(f_i,e_i)}{\text{count}(e_i)} \cdot p_n(f_i|e_i)$

$\phi(f_1)$
$\phi(f_2)$
$\phi(f_3)$
$\phi$(NULL)

Sources of Alignments

The Pharaoh Decoder

- Probabilities at each step include LM and TM
Hypothesis Lattices

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

Pruning

What’s Next?
- Modeling syntax
  - PCFGs and phrase structure
  - Syntactic parsing
  - Grammar induction
  - Syntactic language and translation models
- Speech systems
  - Acoustics
  - Applications