Decoding
The Story So Far...

- Training data (parallel text)
- Learner
- Model
- Decoder

- However, the sky remained clear under the strong north wind.
SCHEDULE

- TODAY
  - word-based translation (conceptual)

- THURSDAY
  - incorporating model scores
  - efficiency considerations
Decoding

- alignment

- \textit{input}: parallel sentences

- \textit{task}: learn a model of translation (probabilities of word translations, fertilities, and distortions)

- decoding

- \textit{input}: foreign-language sentence and a model

- \textit{task}: find the model’s favorite translation
Decoding

- *the process of producing a translation of a sentence*

- Two main problems:
  
  - **modeling** How do we score translations?
  
  - **search** How do we find the model’s preferred translation?
- Decoding stories

  - **Model 1**: *translate each word, scramble them*

  - parameters: $p(e \mid f)$
Das Haus ist klein.

The house is small.
- Decoding stories

  - **Model 2**: translate each word, order them conditioned on absolute position

  - parameters: $p(e \mid f)$

  - new parameters: $a(i \mid j, |E|, |F|)$
Model 2

Lexical translation step

alignment step

natural

ich

of course is the house small

of course the house is small
- Decoding stories

- **Model 3**: model how often each word likes to get translated

- parameters: $p(e \mid f)$, $a(i \mid j, |E|, |F|)$

- new parameters: $n(\phi \mid f)$
MODEL 3

fertility step

NULL insertion step

lexical translation step

distortion step

ich gehe ja nicht zum haus

ich gehe nicht zum zum haus

ich NULL gehe nicht zum zum haus

I do go not to the house

I do not go to the house
- These models are actually quite a bit more complicated than what we actually use in word-based decoding

- Today: you’re the model

- Thursday: we’ll use a real, learned model
- How do we find the translation?

- Easy way: generate all sentences, score them with the model

- Formulate this as stack-based decoding
DECODING (SEARCH)

- Start with a list of hypotheses, containing only the empty hypothesis

- For each stack

  - For each hypothesis

    - For each applicable word

      - Extend the hypothesis with the word

      - Place the new hypothesis on the right stack
Demo

Simple Spanish example (ttable.js)
- Which hypothesis should we extend?

  - *Long hypotheses compete with short ones*

  - *Use multiple stacks for organization*
Demo

Longer Spanish example (ttable-spanish.js)
- There are too many hypotheses!

- *Restrict where the next word can come from.*

- *Monotonic translations*

- *Within n words, n ∈ {1,2,3}*

- *Anywhere*
Activity

http://cs.jhu.edu/~post/mt-class/stack-decoder/index.html

Instructions (10 minutes)
Get in groups, and build out the complete table hypothesis chart using different constraints:
- monotonic
- 1 word distortion
- 2 word distortion
- 3 word distortion
CONCEPTS

- **distortion**: which words are permitted to extend a hypothesis

- for Thursday

  - **pruning**: *histogram* and *beam/threshold*

  - **dynamic programming**

  - *beam* and *threshold* pruning
Demo

German example (from textbook; ttable-german.js)
Types of Errors

- Search errors
  - we didn’t look hard enough (or smart enough)
  - shortcuts we took excluded good hypotheses

- Model errors
  - the solution was impossible under our assumptions
  - the ideal answer cannot be obtained by the model