# Decoding
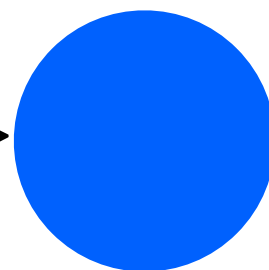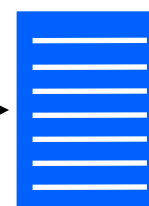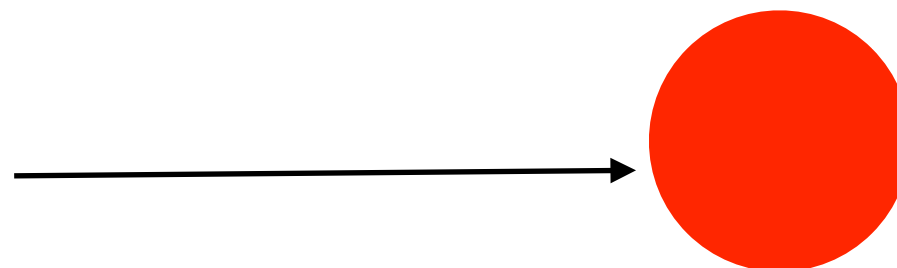
training data
(parallel text)
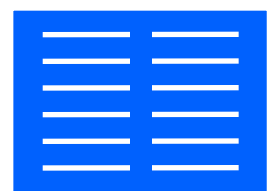
learner

model

联合国 安全 理事会 的

五个 常任 理事 国都

decoder

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

2

# SCHEDULE

- TODAY

    - word-based translation (conceptual)

- THURSDAY

    - incorporating model scores

    - efficiency considerations

# DECODING

- alignment

  - *input:* parallel sentences

  - *task:* learn a model of translation (probabilities of word translations, fertilities, and distortions)

- decoding

  - *input*: foreign-language sentence and a model

  - *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 (MODELING)

- Decoding stories

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

  - parameters: p(e | f)
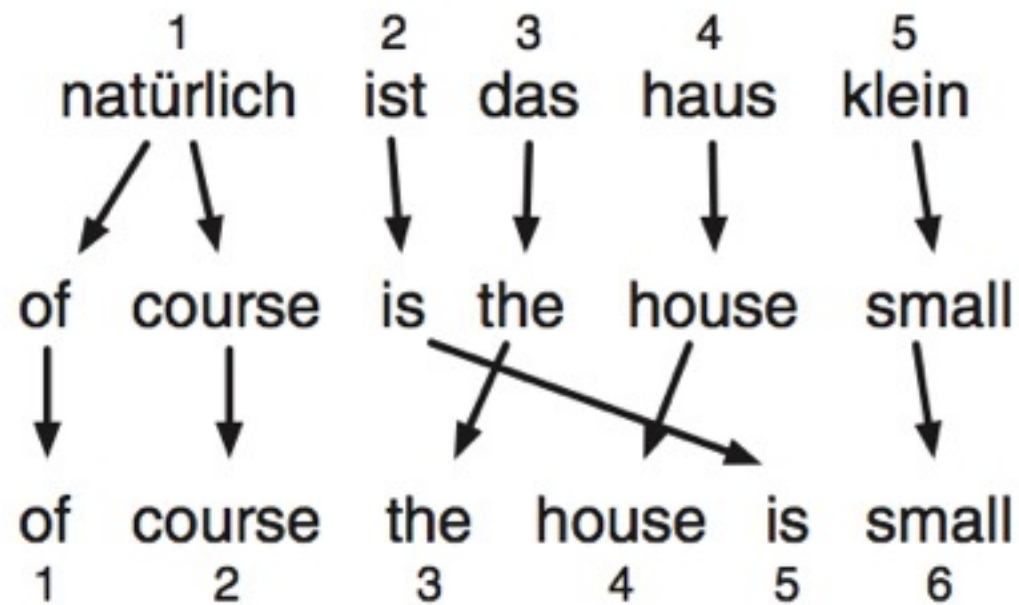
# MODEL 1

# DECODING (MODELING)

- Decoding stories

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

    - parameters: p(e | f)

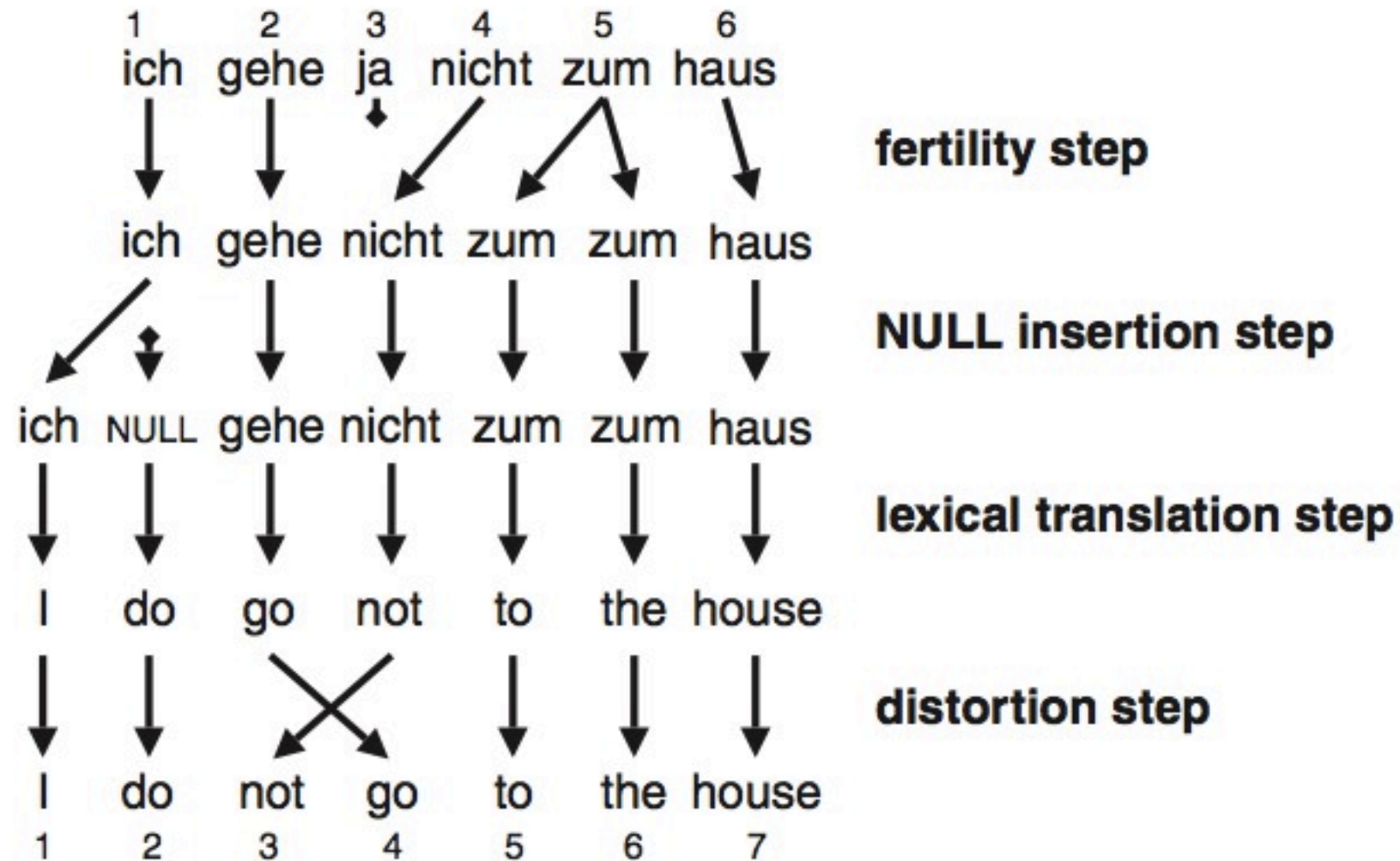    - new parameters: a(i | j, |E|, |F|)

# MODEL 2

# DECODING (MODELING)

- Decoding stories

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

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

    - new parameters: n(φ | f)

# MODEL 3

# DECODING (MODELING)

- 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

# DECODING (SEARCH)

- 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

- At each step:

  - Choose a hypothesis from the list

    - if it's complete, output it and quit

  - Extend it with an untranslated word

  - Place the new hypothesis back on the list

# Demo

Simple Spanish example (ttable.js)

# PROBLEM

- Which hypothesis should we extend?

  - *Long hypotheses compete with short ones*

  - *Use multiple stacks for organization*

# Demo

Longer Spanish example (ttable-spanish.js)

# PROBLEM

- There are too many hypotheses!

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

    - *Monotonic translations*

    - *Within* n *words, n $\in$ {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