WHAT WE HAVE?

FOREIGN/
ENGLISH
TEXT

ENGLISH
TEXT

TRAINING

DECODING

Wednesday, November 21, 12
Translation and Language Models

Foreign --> English MT Model

Foreign/English Text → Translation Model → English Text → Language Model

Training

Decoding

Wednesday, November 21, 12
**Translation and Language Models**

**Foreign --> English MT Model**

- **Foreign/English Text**
- **Translation Model**
- **English Text**
- **Language Model**

**Training**

**Decoding**

English output

澳 洲是与北韩有 邦交的少 数国家之一
Foreign --> English MT Model

Foreign/English Text

Translation Model

English Text

Language Model

Decoder

澳 洲是与北韩有 邦交的少 数国家之一
BRUTE FORCE SEARCH

- Simple minded way to do translation is like what you are doing for homework 3

- Find translations of each word with highest translation score and append them together

- List all possible translations, score them, and pick the best

- Not good: why?
BRUTE FORCE SEARCH

A simple example:

- 16 characters. Let’s say each character has 5 possible translations

- $5^{16} = 152$ billion possible translations

- Phrase-based translation --> different segmentation of the foreign sentence

- Re-ordering. A lot more combinations possible!
We need a strategy to search among all possible translations in an efficient manner.

“Decoder” is the component that does the search.
**Decoder and Search Space**

- **Usually the search space is specified and limited by a grammar and/or the decoder itself**

- **Emphasis is on efficiency while making as few search errors as possible**

![Diagram](image)

- "best" translation
  - cost = 6.2
  - (not reachable)

- Decoder output (English)
  - cost = 7.1

- Another translation
  - cost = 6.8
  - (search error)
GENERAL SEARCH STRATEGY

- Create and Score “Partial Hypothesis”

- Each Partial Hyp is the translation of certain words of foreign language

- Group Partial Hyps

- In each group keep only the most promising Partial Hyps

- Extend Partial Hyps
Decoders:

- Phrase Decoder
- Hierarchical Decoder (Chart)
- Left-to-Right hierarchical
- Your own decoding strategy...
Phrasal-based Decoder

Other names: Beam Decoder, Stack Decoder, ...

Left-to-right production of translation

But can move around foreign sentence

The moves are not totally arbitrary.

Why?
Decoding Process

- Build translation left to right
  - select foreign words to be translated
Decoding Process

- Build translation left to right
  - select foreign words to be translated
  - find English phrase translation
  - add English phrase to end of partial translation
Decoding Process

- Build translation left to right
  - select foreign words to be translated
  - find English phrase translation
  - add English phrase to end of partial translation
  - mark foreign words as translated
Decoding Process

- One to many translation
Decoding Process

- Many to one translation
Pharaoh: Beam Search Decoder for Phrase-Based Statistical Machine Translation

Decoding Process

- Many to one translation

Maria no dio una bofetada a la bruja verde

Mary did not slap the verde
Decoding Process

- Reordering

Maria no dio una bofetada a la bruja verde

Mary did not slap the green
Decoding Process

- Translation finished
Translation Options

<table>
<thead>
<tr>
<th>Maria</th>
<th>no</th>
<th>dio</th>
<th>una</th>
<th>bofetada</th>
<th>a</th>
<th>la</th>
<th>bruja</th>
<th>verde</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mary</td>
<td>not</td>
<td>give</td>
<td>a</td>
<td>slap</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>did not</td>
<td></td>
<td>a slap</td>
<td>by</td>
<td>the witch</td>
<td>green</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>no</td>
<td></td>
<td>slap</td>
<td>to the</td>
<td>green witch</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>did not give</td>
<td></td>
<td></td>
<td>to</td>
<td>the</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Look up possible phrase translations
  - many different ways to segment words into phrases
  - many different ways to translate each phrase
Hypothesis Expansion

<table>
<thead>
<tr>
<th>Maria</th>
<th>no</th>
<th>dio</th>
<th>una</th>
<th>bofetada</th>
<th>a</th>
<th>la</th>
<th>bruja</th>
<th>verde</th>
</tr>
</thead>
</table>

- Mary ___ not ___ give ___ a ___ slap ___ to ___ the ___ witch ___ green
- did not ___ a slap ___ by ___ green witch
- no ___ slap ___ to the
- did not give ___ to ___ the
- slap ___ the witch

e: __________
f: __________
p: 1

● Start with null hypothesis
  - e: no English words
  - f: no foreign words covered
  - p: probability 1
### Hypothesis Expansion

<table>
<thead>
<tr>
<th>Maria</th>
<th>no</th>
<th>dio</th>
<th>una</th>
<th>bofetada</th>
<th>a</th>
<th>la</th>
<th>bruja</th>
<th>verde</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Pick translation option**
- **Create hypothesis**
  - e: add English phrase *Mary*
  - f: first foreign word covered
  - p: probability 0.534
Hypothesis Expansion

<table>
<thead>
<tr>
<th>Maria</th>
<th>no</th>
<th>dio</th>
<th>una</th>
<th>bofetada</th>
<th>a</th>
<th>la</th>
<th>bruja</th>
<th>verde</th>
</tr>
</thead>
</table>

Mary did not give a slap to the witch

Did not give a slap to the green witch

---

**Add another hypothesis**
Hypothesis Expansion

- Further hypothesis expansion
Hypothesis Expansion

- Mary did not give a slap to the witch by green witch
- no slap to the green witch
- did not give
- slap to the witch

... until all foreign words covered
- find best hypothesis that covers all foreign words
- backtrack to read off translation

Philipp Koehn, Massachusetts Institute of Technology

Wednesday, November 21, 12
Hypothesis Expansion

<table>
<thead>
<tr>
<th>Maria</th>
<th>no</th>
<th>dio</th>
<th>una</th>
<th>bofetada</th>
<th>a</th>
<th>la</th>
<th>bruja</th>
<th>verde</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mary</td>
<td>not</td>
<td>give</td>
<td>a</td>
<td>slap</td>
<td>to</td>
<td>the</td>
<td>witch</td>
<td>green</td>
</tr>
<tr>
<td>did not</td>
<td></td>
<td>a slap</td>
<td></td>
<td>by</td>
<td>to the</td>
<td>green witch</td>
<td></td>
<td></td>
</tr>
<tr>
<td>no</td>
<td></td>
<td>slap</td>
<td></td>
<td>to</td>
<td>the</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>did not give</td>
<td></td>
<td></td>
<td></td>
<td>the</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>slap</td>
<td></td>
<td></td>
<td></td>
<td>the witch</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Adding more hypothesis

⇒ Explosion of search space
Restricting the search space

Search space is too big.

Use the two following restrictions on movements:

1. **Skip**: how many source words can we temporarily skip during translation

2. **Window Width**: how far to the right can we go before we are forced to translate a skipped word

WW=9

S=5
Explosion of Search Space

- Number of hypotheses is exponential with respect to sentence length

⇒ Decoding is NP-complete [Knight, 1999]
⇒ Need to reduce search space
  - risk free: hypothesis recombination
  - risky: histogram/threshold pruning
Hypothesis Recombination

- Different paths to the same partial translation
Hypothesis Recombination

- Different paths to the same partial translation
  - Combine paths
    - drop weaker hypothesis
    - keep pointer from worse path
Recombined hypotheses do not have to match completely

No matter what is added, weaker path can be dropped, if:
- last two English words match (matters for language model)
- foreign word coverage vectors match (effects future path)
Hypothesis Recombination

- Recombined hypotheses do not have to match completely.
- No matter what is added, weaker path can be dropped, if:
  - last two English words match (matters for language model)
  - foreign word coverage vectors match (effects future path)

⇒ Combine paths
Pruning

- Hypothesis recombination is not sufficient

⇒ Heuristically discard weak hypotheses

- Organize Hypothesis in stacks, e.g. by
  - same foreign words covered
  - same number of foreign words covered (Pharaoh does this)
  - same number of English words produced

- Compare hypotheses in stacks, discard bad ones
  - histogram pruning: keep top \( n \) hypotheses in each stack (e.g., \( n=100 \))
  - threshold pruning: keep hypotheses that are at most \( \alpha \) times the cost of best hypothesis in stack (e.g., \( \alpha = 0.001 \))
Comparing Hypotheses

- Comparing hypotheses with same number of foreign words covered

Maria no dio una bofetada a la bruja verde

\[
\begin{align*}
\text{e: Mary did not} \\
\text{f: *-*-*-*-*} \\
\text{p: 0.154}
\end{align*}
\]

\[
\begin{align*}
\text{e: the} \\
\text{f: ----*-*---} \\
\text{p: 0.354}
\end{align*}
\]

- Better partial translation
- Covers easier part --> lower cost

- Hypothesis that covers easy part of sentence is preferred

\[ \Rightarrow \text{Need to consider future cost} \]
Future Cost Estimation

- Estimate cost to translate remaining part of input

- **Step 1: find cheapest translation options**
  - find cheapest translation option for each input span
  - compute translation model cost
  - estimate language model cost (no prior context)
  - ignore reordering model cost

- **Step 2: compute cheapest cost**
  - for each contiguous span:
    - find cheapest sequence of translation options

- **Precompute and lookup**
  - precompute future cost for each contiguous span
  - future cost for any coverage vector:
    - sum of cost of each contiguous span of uncovered words
  - no expensive computation during run time
Outline

- Phrase-Based Statistical MT
- Beam Search Decoding
- Experiments
- Advanced Features
Experiments

- Decoder has to be evaluated in terms of search errors
  - translation errors not due to search errors are a challenge to the translation model
  - do not rely on search errors for good translation quality!

- Experimental setup
  - German to English
  - Europarl training corpus (30 million words)
  - 1500 sentence test corpus (avg. length 28.9 words)
  - 3 Ghz Linux machine, needs 512 MB RAM
  - Focus: illustrate trade-off speed / search errors

- Not measuring true search error
  - it is not tractable to find truly best translation
  - relative to best translation found with high beam and different settings
Threshold Pruning

<table>
<thead>
<tr>
<th>Threshold</th>
<th>0.0001</th>
<th>0.001</th>
<th>0.01</th>
<th>0.05</th>
<th>0.08</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time per Sentence</td>
<td>149 sec</td>
<td>119 sec</td>
<td>70 sec</td>
<td>27 sec</td>
<td>18 sec</td>
</tr>
<tr>
<td>Search Errors</td>
<td>-</td>
<td>+0%</td>
<td>+0%</td>
<td>+0%</td>
<td>+0%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Threshold</th>
<th>0.1</th>
<th>0.15</th>
<th>0.2</th>
<th>0.3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time per Sentence</td>
<td>15 sec</td>
<td>13 sec</td>
<td>10 sec</td>
<td>7 sec</td>
</tr>
<tr>
<td>Search Errors</td>
<td>+1%</td>
<td>+3%</td>
<td>+6%</td>
<td>+12%</td>
</tr>
</tbody>
</table>

- Low ratio of search errors for threshold $\alpha \leq 0.1$

- Results depend on weights for models
## Histogram Pruning

<table>
<thead>
<tr>
<th>Beam Size</th>
<th>1000</th>
<th>200</th>
<th>100</th>
<th>50</th>
<th>20</th>
<th>10</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>15s</td>
<td>15s</td>
<td>14s</td>
<td>10s</td>
<td>9s</td>
<td>9s</td>
<td>7s</td>
</tr>
<tr>
<td>Search Errors</td>
<td>+1%</td>
<td>+1%</td>
<td>+2%</td>
<td>+4%</td>
<td>+8%</td>
<td>+20%</td>
<td>+35%</td>
</tr>
</tbody>
</table>

- Low ratio of search errors for beam size $n \geq 200$
### Translation Table Entries per Input Phrase

<table>
<thead>
<tr>
<th>T-Table Limit</th>
<th>1000</th>
<th>500</th>
<th>200</th>
<th>100</th>
<th>50</th>
<th>20</th>
<th>10</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Time</strong></td>
<td>15.0s</td>
<td>7.6s</td>
<td>3.8s</td>
<td>1.9s</td>
<td>0.9s</td>
<td>0.4s</td>
<td>0.2s</td>
<td>0.1s</td>
</tr>
<tr>
<td><strong>Search Errors</strong></td>
<td>+1%</td>
<td>+1%</td>
<td>+1%</td>
<td>+1%</td>
<td>+1%</td>
<td>+2%</td>
<td>+7%</td>
<td>+18%</td>
</tr>
</tbody>
</table>

- Low ratio of search errors for limit of $\geq 50$ entries in the translation table for each source language phrase
- About 1 second per sentence (30 words per second)
- Your mileage may vary
Hierarchical Decoder
Australia is with North Korea have diplomatic relations that few countries one of...
Australia is with North Korea have diplomatic relations that few countries one of
Australia is with North Korea is diplomatic relations one of the few countries
Australia is with North Korea is diplomatic relations with one of the few countries.

WHAT’S WRONG HERE?
Phrase-based vs. Hierarchical

- Phrase-based decoder is good at capturing local re-orderings
- Hierarchical decoding can capture re-orderings among phrases themselves
- Useful for capturing long-distance movements
Hierarchical Translation

澳洲 是 与北 韩 有 邦交 的 少数 国家 之一
Australia is North Korea's few diplomatic relations with a few countries.
Australia is with North Korea and has diplomatic relations with few countries.

Hierarchical Translation
Australia is with North Korea having diplomatic relations with few countries.
Australia is have diplomatic relations with North Korea

few countries
Hierarchical Translation

Australia is few countries that have diplomatic relations with North Korea.

The few countries that have diplomatic relations with North Korea.

Australia
Australia is one of the few countries that have diplomatic relations with North Korea.
Hierarchical Translation

Australia is one of the few countries that have diplomatic relations with North Korea.

Hierarchical Translation problem we saw before is solved!
Hierarchical Rules

Extracted from phrases
(you know phrase extraction from previous lecture)

与中国有外交谈判 ➔ have diplomatic talks with China
外交谈判 ➔ diplomatic talks
中国 ➔ China
Hierarchical Rules

Extracted from phrases
(you know phrase extraction from previous lecture)

与中国有外交谈判 → have diplomatic talks with China
外交谈判 → diplomatic talks
中国 → China

X: 与 [X1] 有 [X2], have [X2] with [X1]
<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>AUSTRALIA</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>IS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>NORTH KOREA</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>DIPLOMATIC RELATIONS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>FEW COUNTRIES</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Australia is one of the few countries with diplomatic relations with North Korea.**
Australia is one of the few countries that have diplomatic relations with North Korea.
Australia is one of the few countries that have diplomatic relations with North Korea.
Australia is one of the few countries that have diplomatic relations with North Korea.
Australia is North Korea diplomatic relations few countries.

X: [X1] [X2], [X1] [X2] "glue rule"
Australia is one of the few countries that have diplomatic relations with North Korea.
Chart Decoding: Complexity

\[ n \frac{(n+1)}{2} h^2 r \]

- \( n \): sentence length (in 10s)
- \( h \): # kept hyps in each cell (in 100s)
- \( r \): # rules in each cell (in 10s)

\( h^2 \) contributes the most to the complexity
**Problem:** Find the best 5 combined hyps without trying all 25 possible combinations.

<table>
<thead>
<tr>
<th></th>
<th>h₁</th>
<th>h₂</th>
<th>h₃</th>
<th>h₄</th>
<th>h₅</th>
</tr>
</thead>
<tbody>
<tr>
<td>g₁</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>g₂</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>g₃</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>g₄</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>g₅</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

X: 与 [X1] 有 [X2], have [X2] with [X1]
The hyps in the lower cell are ordered

**But:** \( h_1 < h_2, g_1 < g_2 \not\Rightarrow (h_1 + g_1) < (h_2 + g_2) \)

**We need a strategy to explore the most promising combinations**
Chart Decoding: Cube Pruning

**Why**  
$h_1 < h_2, g_1 < g_2 \implies (h_1 + g_1) < (h_2 + g_2)$  

**Main reason is language model (or any context-sensitive model)**

**P(is) < P(was) doesn’t necessarily mean P(Australia is) < P(Australia was)**

\[
P(\text{Australia is}) = P(\text{Australia}) + P(\text{is} | \text{Australia})
\]

\[
P(\text{Australia was}) = P(\text{Australia}) + P(\text{was} | \text{Australia})
\]

**Context changes. True for all language models orders except unigram**
Chart Decoding: Cube Pruning

- Start with the top-right corner, build the new hyp, and add the two neighboring combinations to the priority list.

\[
\begin{array}{ccccc}
  h_1 & h_2 & h_3 & h_4 & h_5 \\
  g_1 & & C_{1,2} & & \\
  g_2 & & C_{2,1} & & \\
  g_3 & & & & \\
  g_4 & & & & \\
  g_5 & & & & \\
\end{array}
\]

X: 与 [X1] 有 [X2], have [X2] with [X1]
Chart Decoding: Cube Pruning

- Pick the lowest cost item in the priority queue, build hypothesis and add neighbors
- Priority queue ordered by approximate cost: \( c_{x,y} = h_x + g_y \)

<table>
<thead>
<tr>
<th></th>
<th>( h_1 )</th>
<th>( h_2 )</th>
<th>( h_3 )</th>
<th>( h_4 )</th>
<th>( h_5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( g_1 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( g_2 )</td>
<td>✔️</td>
<td>( C_{1,2} )</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( g_3 )</td>
<td>✔️</td>
<td>✔️</td>
<td>( C_{2,2} )</td>
<td></td>
<td></td>
</tr>
<tr>
<td>( g_4 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( g_5 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\( X: 与 [X1] 有 [X2] , have [X2] with [X1] \)
## Chart Decoding: Cube Pruning

<table>
<thead>
<tr>
<th>g1</th>
<th>g2</th>
<th>g3</th>
<th>g4</th>
<th>g5</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="green" alt="g1" /></td>
<td><img src="green" alt="g2" /></td>
<td><img src="C3,2" alt="g3" /></td>
<td><img src="null" alt="g4" /></td>
<td><img src="null" alt="g5" /></td>
</tr>
</tbody>
</table>

- **h1**: ![g1](green)
- **h2**: ![g2](green)
- **h3**: ![g3](C3,2)
- **h4**: ![g4](null)
- **h5**: ![g5](null)

**X:** 与 [X1] 有 [X2], have [X2] with [X1]
Chart Decoding: Cube Pruning

X: 与 [X1] 有 [X2], have [X2] with [X1]
### Chart Decoding: Cube Pruning

<table>
<thead>
<tr>
<th></th>
<th>h1</th>
<th>h2</th>
<th>h3</th>
<th>h4</th>
<th>h5</th>
</tr>
</thead>
<tbody>
<tr>
<td>g1</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>C₁,₄</td>
<td></td>
</tr>
<tr>
<td>g2</td>
<td>✓</td>
<td>C₂,₂</td>
<td>✓</td>
<td>C₂,₄</td>
<td></td>
</tr>
<tr>
<td>g₃</td>
<td></td>
<td>C₃,₂</td>
<td></td>
<td>C₃,₃</td>
<td></td>
</tr>
<tr>
<td>g₄</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>g₅</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

X: 与 [X1] 有 [X2], have [X2] with [X1]
Stopping criteria:

- Stop when a pre-defined number of hyphs are generated (green checkmarks)

- Stop when the cost difference between the most recent hyp and the best one so far is higher than a pre-defined value
Dynamic Programming

Combine two hyps if their completion costs are the same

ie. however we extend one, we can extend the other in the same exact way, and with the same exact additional cost

Define “states” based on the completion concept

Combine hyps that share the same state
Chart Decoding: Hyp Combination

- Only compare and define hyps in the same cell (same coverage of foreign words)
- States are defined based on language model (or any additional context-sensitive models)
- We can append to a hyp from either left or right
- Language model state is defined as the \((n-1)\) left-most and \((n-1)\) right-most words

Few countries that have diplomatic relations with North Korea
Cost of words in the middle are not going to change

\[ P(\text{that}) = P(\text{that} \mid \text{few countries}) \]

Continuation costs depend on the last 2 words only

\[ P( . \mid \text{North Korea}) \]

Does not depend on what words were there before “North Korea”
Among the hyphs in the same cell, sharing the same (LM) states, keep only the best (lowest-cost) one.

No need to explore the expansion of the higher-cost hyphs with the same state.

The higher-cost hyphs can be kept and used for building $N$-best lists.
Search Space:

Phrase-bases vs. Hierarchical
Coverage Vectors: N=6, all possible permutations
Coverage Vectors: N=6, all possible permutations

1 2 3 4 5 6 1
Coverage Vectors: N=6, all possible permutations
Coverage Vectors: N=6, all possible permutations
Coverage Vectors: N=6, all possible permutations
Coverage Vectors: N=6, all possible permutations
Coverage Vectors: N=6, all possible permutations
Coverage Vectors: N=6, all possible permutations

states= 64
$N=6$, Phrase-based, $S=2$, $WW=3$
N=6, Phrase-based, S=2, WW=3
N=6, Phrase-based, S=2, WW=3
N=6, Phrase-based, S=2, WW=3

states=29 out of 64
**Chart-based Decoder**

<table>
<thead>
<tr>
<th>CARDINALITY</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SOURCE</th>
<th>POSITION</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Chart-based Decoder

Cardinality

Source Position

1 2 3 4 5 6
1 2 3 4 5 6
1 2 3 4 5 6
1 2 3 4 5 6
1 2 3 4 5 6
1 2 3 4 5 6
Chart-based Decoder

Cardinality

Source Position

Wednesday, November 21, 12
Chart-based Decoder

Cardinality

Source Position

Wednesday, November 21, 12
Chart-based Decoder

Cardinality

Source Position

Wednesday, November 21, 12
Chart-based Decoder

Cardinality

Source

Position

Wednesday, November 21, 12
ITG Constraint

\begin{align*}
[[[1,2],3],4] & \quad [[[3,2],1],4] & \quad [[[2,3],4],1] \\
[2],[4],[1],[3] & \quad [3],[1],[4],[2]
\end{align*}
Simplified Hierarchical

Inversion Transduction Grammars (ITG)

Wu, 1997

X: [X1] [X2] , [X1] [X2]  
   “glue rule”

X: [X1] [X2] , [X2] [X1]  
   “swap rule”
N=6: Hierarchical (ITG)
N=6: Hierarchical (ITG)
N=6: Hierarchical (ITG)
N=6: Hierarchical (ITG)
N=6: Hierarchical (ITG)

states = 21
phrase-based = 29, n=11
## Chart Decoder Search Space:

<table>
<thead>
<tr>
<th></th>
<th>Sentence Length = n</th>
<th>No Constraints</th>
<th>Chart ITG Constraints</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>States</strong></td>
<td>Unique Coverage Vectors</td>
<td>$2^n$</td>
<td>$\frac{n(n+1)}{2}$</td>
</tr>
<tr>
<td><strong>Edges</strong></td>
<td>Extensions</td>
<td>$\sum_{i=0}^{n} \text{Choose}(n,k) 	imes (n-i)$</td>
<td>$\frac{2n+n^3}{3}$</td>
</tr>
<tr>
<td><strong>Paths</strong></td>
<td>Source Order Permutations</td>
<td>$n!$</td>
<td>?</td>
</tr>
</tbody>
</table>
## Comparison

<table>
<thead>
<tr>
<th>N=6</th>
<th>N=10</th>
</tr>
</thead>
<tbody>
<tr>
<td>ALL</td>
<td>ITG</td>
</tr>
<tr>
<td><strong>States</strong></td>
<td>64</td>
</tr>
<tr>
<td><strong>Edges</strong></td>
<td>192</td>
</tr>
<tr>
<td><strong>Paths</strong></td>
<td>720</td>
</tr>
</tbody>
</table>
Search space comparison

- Phrase-based
- Monotone
- Hierarchical

All possible translations