Machine Translation: Searching for Good Translations

Introduction to Natural Language Processing
Computer Science 585—Fall 2009
University of Massachusetts Amherst

David Smith
Search

What’s the best translation (under our model)?
Search

• Even if we know the right words in a translation, there are \( n! \) permutations.
  
  \[
  10! = 3,626,800 \quad 20! \approx 2.43 \times 10^{18} \quad 30! \approx 2.65 \times 10^{32}
  \]

• We want the translation that gets the highest score under our model
  – Or the best \( k \) translations
  – Or a random sample from the model’s distribution

• But **not** in \( n! \) time!
Deshalb haben wir allen Grund, die Umwelt in die Agrarpolitik zu integrieren.
Search in Phrase Models

Deshalb haben wir allen Grund, die Umwelt in die Agrarpolitik zu integrieren.

That is why we have

Translate in target language order to ease language modeling.
Search in Phrase Models

Deshalb haben wir allen Grund, die Umwelt in die Agrarpolitik zu integrieren

That is why we have every reason

*Translate in target language order to ease language modeling.*
Deshalb haben wir allen Grund, die Umwelt in die Agrarpolitik zu integrieren,

That is why we have every reason to
Search in Phrase Models

Deshalb haben wir allen Grund, die Umwelt in die Agrarpolitik zu integrieren

That is why we have every reason to integrate

Translate in target language order to ease language modeling.
That is why we have every reason to integrate the environment into agricultural policy.

*Translate in target language order to ease language modeling.*
That is why we have every reason to integrate the environment in the agrarpolitik to ease language modeling.
That is why we have every reason to integrate the environment in the Agrarpolitik zu integrieren.
Deshalb haben wir allen Grund, die Umwelt in die Agrarpolitik zu integrieren

That is why we have every reason to integrate the environment in the agricultural policy.

Translate in target language order to ease language modeling.
That is why we have every reason to integrate the environment in the agricultural policy.
That is why we have every reason to integrate the environment in the agricultural policy.
That is why we have every reason to integrate the environment in the agricultural policy.
Deshalb haben wir allen Grund, die Umwelt in die Agrarpolitik zu integrieren.

Therefore, we have every reason, the environment in the agricultural policy, to integrate.

That is why we have all reason, which environment in the agricultural policy, parliament.

We have therefore us all the reason of the environment into the agricultural policy, successfully integrated.

Hence, we every reason to make environmental on the cap, be woven together.

We have therefore everyone grounds for taking the environment to the agricultural policy is on parliament.

So, we all of cause which environment, to the cap, for incorporated.

Hence our any why that outside at agricultural policy too woven together.

Therefore, it of all reason for, the completion into that agricultural policy be.
Deshalb haben wir allen Grund, die Umwelt in die Agrarpolitik zu integrieren.

<table>
<thead>
<tr>
<th>Deshalb</th>
<th>haben</th>
<th>wir</th>
<th>allen</th>
<th>Grund</th>
<th>,</th>
<th>die</th>
<th>Umwelt</th>
<th>in</th>
<th>die</th>
<th>Agrarpolitik</th>
<th>zu</th>
<th>integrieren</th>
</tr>
</thead>
<tbody>
<tr>
<td>that is why we have</td>
<td>every reason</td>
<td>the environment</td>
<td>in</td>
<td>the</td>
<td>agricultural policy</td>
<td>to</td>
<td>integrate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>therefore have we</td>
<td>every reason</td>
<td>the environment</td>
<td>in</td>
<td>the</td>
<td>agricultural policy</td>
<td>,</td>
<td>to integrate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>that is why we have</td>
<td>all</td>
<td>reason</td>
<td>,</td>
<td>which</td>
<td>environment in</td>
<td>agricultural policy</td>
<td>parliament</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>have therefore us</td>
<td>all the</td>
<td>reason</td>
<td>of the</td>
<td>environment into</td>
<td>the agricultural policy</td>
<td>successfully integrated</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hence , we</td>
<td>every reason</td>
<td>to make</td>
<td>environmental</td>
<td>on</td>
<td>the cap</td>
<td>be woven together</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>we have therefore everyone</td>
<td>grounds for taking the</td>
<td>the environment</td>
<td>to the</td>
<td>agricultural policy</td>
<td>is</td>
<td>on</td>
<td>parliament</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>so , we</td>
<td>all of</td>
<td>cause</td>
<td>which</td>
<td>environment ,</td>
<td>to</td>
<td>the cap ,</td>
<td>for</td>
<td>incorporated</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hence our</td>
<td>any</td>
<td>why</td>
<td>that</td>
<td>outside</td>
<td>at</td>
<td>agricultural policy</td>
<td>too</td>
<td>woven together</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>therefore ,</td>
<td>it</td>
<td>of all</td>
<td>reason for</td>
<td>,</td>
<td>the completion</td>
<td>into</td>
<td>that agricultural policy</td>
<td>be</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

And many, many more…even before reordering
### Search in Phrase Models

<table>
<thead>
<tr>
<th>Deshalb</th>
<th>haben</th>
<th>wir</th>
<th>allen</th>
<th>Grund</th>
<th>,</th>
<th>die</th>
<th>Umwelt</th>
<th>in</th>
<th>die</th>
<th>Agrarpolitis</th>
<th>zu</th>
<th>integrieren</th>
</tr>
</thead>
<tbody>
<tr>
<td>that is why we have</td>
<td>every reason</td>
<td>the environment</td>
<td>in</td>
<td>the</td>
<td>agricultural policy</td>
<td>to</td>
<td>integrate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>therefore</td>
<td>have</td>
<td>we</td>
<td>every reason</td>
<td>the</td>
<td>environment</td>
<td>in the</td>
<td>agricultural policy ,</td>
<td>to integrate</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>that is why</td>
<td>we have</td>
<td>all</td>
<td>reason</td>
<td>, which</td>
<td>environment in</td>
<td>agricultural policy</td>
<td>parliament</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>have therefore</td>
<td>us</td>
<td>all the</td>
<td>reason</td>
<td>of the</td>
<td>environment into</td>
<td>the agricultural policy</td>
<td>successfully integrated</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hence</td>
<td>, we</td>
<td>every</td>
<td>reason to make</td>
<td>environmental</td>
<td>on</td>
<td>the cap</td>
<td>be woven together</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>we have therefore</td>
<td>everyone</td>
<td>grounds for taking the</td>
<td>the environment</td>
<td>to the</td>
<td>agricultural policy is</td>
<td>on</td>
<td>parliament</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>so</td>
<td>, we</td>
<td>all of</td>
<td>cause</td>
<td>which</td>
<td>environment ,</td>
<td>to</td>
<td>the cap ,</td>
<td>for</td>
<td>incorporated</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hence our</td>
<td>any</td>
<td>why</td>
<td>that</td>
<td>outside</td>
<td>at</td>
<td>agricultural policy</td>
<td>too</td>
<td>woven together</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>therefore ,</td>
<td>it</td>
<td>of all</td>
<td>reason for</td>
<td>, the</td>
<td>completion</td>
<td>into</td>
<td>that agricultural policy</td>
<td>be</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*And many, many more…even before reordering*
<table>
<thead>
<tr>
<th>German</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deshalb</td>
<td>that is why we have</td>
</tr>
<tr>
<td>haben</td>
<td>every reason</td>
</tr>
<tr>
<td>wir</td>
<td>the environment</td>
</tr>
<tr>
<td>allen</td>
<td>in the</td>
</tr>
<tr>
<td>Grund ,</td>
<td>the agricultural policy to integrate</td>
</tr>
<tr>
<td>die</td>
<td></td>
</tr>
<tr>
<td>Umwelt</td>
<td></td>
</tr>
<tr>
<td>in die</td>
<td></td>
</tr>
<tr>
<td>Agrarpolitik</td>
<td>zu integrieren</td>
</tr>
<tr>
<td>zu</td>
<td></td>
</tr>
<tr>
<td>integieren</td>
<td></td>
</tr>
<tr>
<td></td>
<td>hence every reason environmental on the cap be woven together</td>
</tr>
<tr>
<td></td>
<td>we have all the reason of the environment into the agricultural policy successfully integrated</td>
</tr>
<tr>
<td></td>
<td>have therefore us all the reason of the environment into the agricultural policy parliament</td>
</tr>
<tr>
<td></td>
<td>we have therefore everyone grounds for taking the environment to the agricultural policy on parliament</td>
</tr>
<tr>
<td></td>
<td>so , we all of cause which environment , to the cap , for incorporated</td>
</tr>
<tr>
<td></td>
<td>hence our any why that outside at agricultural policy too woven together</td>
</tr>
<tr>
<td></td>
<td>therefore , it of all reason for , the completion into that agricultural policy be</td>
</tr>
</tbody>
</table>

And many, many more…even before reordering
<table>
<thead>
<tr>
<th>Deshalb</th>
<th>haben</th>
<th>wir</th>
<th>allen</th>
<th>Grund</th>
<th>die</th>
<th>Umwelt</th>
<th>in</th>
<th>die</th>
<th>Agrarpolitik</th>
<th>zu</th>
<th>integrieren</th>
</tr>
</thead>
<tbody>
<tr>
<td>that is why we have</td>
<td>every reason</td>
<td>the environment</td>
<td>in</td>
<td>the</td>
<td>agricultural policy</td>
<td>to</td>
<td>integrate</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>therefore</td>
<td>have</td>
<td>we</td>
<td>every reason</td>
<td>the</td>
<td>environment</td>
<td>in the</td>
<td>agricultural policy</td>
<td>, to integrate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>that is why</td>
<td>we have</td>
<td>all</td>
<td>reason</td>
<td>, which</td>
<td>environment in</td>
<td>agricultural policy</td>
<td>parliament</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>have therefore</td>
<td>us</td>
<td>all the</td>
<td>reason</td>
<td>of the</td>
<td>environment into</td>
<td>the agricultural policy</td>
<td>successfully integrated</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hence</td>
<td>, we</td>
<td>every</td>
<td>reason to make</td>
<td>environmental</td>
<td>on</td>
<td>the cap</td>
<td>be woven together</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>we have therefore</td>
<td>everyone</td>
<td>grounds for taking the</td>
<td>the environment</td>
<td>to the</td>
<td>agricultural policy is</td>
<td>on</td>
<td>parliament</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>so</td>
<td>, we</td>
<td>all of</td>
<td>cause</td>
<td>which</td>
<td>environment</td>
<td>to</td>
<td>the cap</td>
<td>for incorporated</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hence our</td>
<td>any</td>
<td>why</td>
<td>that</td>
<td>outside</td>
<td>at</td>
<td>agricultural policy</td>
<td>too</td>
<td>woven together</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>therefore</td>
<td>, it</td>
<td>of all</td>
<td>reason for</td>
<td>, the</td>
<td>completion</td>
<td>into</td>
<td>that agricultural policy</td>
<td>be</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
“Stack Decoding”

Deshalb haben wir allen Grund, die Umwelt in die Agrarpolitik zu integrieren.
“Stack Decoding”

Deshalb haben wir allen Grund, die Umwelt in die Agrarpolitik zu integrieren.
Deshalb haben wir allen Grund, die Umwelt in die Agrarpolitik zu integrieren.
“Stack Decoding”

Deshalb haben wir allen Grund, die Umwelt in die Agrarpolitik zu integrieren.

hence
we
have
“Stack Decoding”

Deshalb haben wir allen Grund, die Umwelt in die Agrarpolitik zu integrieren.

- hence
- we
- have
- in
Deshalb haben wir allen Grund, die Umwelt in die Agrarpolitik zu integrieren.
Deshalb haben wir allen Grund, die Umwelt in die Agrarpolitik zu integrieren.
“Stack Decoding”

Deshalb haben wir allen Grund, die Umwelt in die Agrarpolitik zu integrieren.
“Stack Decoding”

Deshalb haben wir allen Grund, die Umwelt in die Agrarpolitik zu integrieren.
“Stack Decoding”

Deshalb haben wir allen Grund, die Umwelt in die Agrarpolitik zu integrieren.

hence, hence we have therefore

we have

have

in the environment

the
“Stack Decoding”

Deshalb haben wir allen Grund, die Umwelt in die Agrarpolitik zu integrieren.

hence
hence we
we
we have
have
in
the environment
Deshalb haben wir allen Grund, die Umwelt in die Agrarpolitik zu integrieren.

We could declare these equivalent.
“Stack Decoding”

Deshalb haben wir allen Grund, die Umwelt in die Agrarpolitik zu integrieren.

We could declare these equivalent.

etc., u.s.w., until all source words are covered.

The environment
Search in Phrase Models

• Many ways of segmenting source
• Many ways of translating each segment
• *Restrict* model class: phrases >, e.g., 7 words, no long-distance reordering
• *Recombine* equivalent hypotheses
• *Prune* away unpromising partial translations or we’ll run out of space and/or run too long
  – How to compare partial translations?
  – Some start with easy stuff: “in”, “das”, ...
  – Some with hard stuff: “Agrarpolitik”, “Entscheidungsproblem”, …
Hypothesis Recombination

• Different paths to the same partial translation

\[
\begin{align*}
p &= 1 \\
\text{Mary} &\quad p &= 0.534 \\
\text{did not} &\quad p &= 0.164 \quad \text{give} \\
\text{did not give} &\quad p &= 0.092 \quad \text{give} \\
&\quad p &= 0.044
\end{align*}
\]
Hypothesis Recombination

- Different paths to the same partial translation
- Combine paths
  - Drop weaker path
  - Keep backpointer to weaker path (for lattice or n-best generation)
Hypothesis Recombination

• Recombined hypotheses do not have to match completely
• Weaker path can be dropped if
  – Last n target words match (for n+1-gram lang. model)
  – Source coverage vectors match (same best future)

\[ p=0.092 \quad \text{did not give} \]
\[ p=0.534 \quad \text{did not give} \]
\[ p=0.164 \quad \text{give} \]
Hypothesis Recombination

• Combining partially matching hypotheses

![Diagram showing hypothesis recombination with probabilities and labels.]
Pruning

• Hypothesis recombination is not sufficient
  Heuristically discard weak hypotheses early

• Organize Hypothesis in stacks, e.g. by
  – same foreign words covered
  – same number of foreign words covered
  – 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 $t$ times the cost of
    best hypothesis in stack (e.g., $t = 0.001$)
Word Lattice Generation

- Search graph can be easily converted into a word lattice
  - can be further mined for n-best lists
    enables reranking approaches
    enables discriminative training

Philipp Koehn  JHU SS  6 July 2006
• Organization of hypothesis into stacks
  
  – here: based on *number of foreign words* translated
  – during translation all hypotheses from one stack are expanded
  – expanded Hypotheses are placed into stacks
Limits on Reordering

• Reordering may be limited
  – **Monotone** Translation: No reordering at all
  – Only phrase movements of at most $n$ words

• Reordering limits *speed* up search (polynomial instead of exponential)

• Current reordering models are weak, so limits *improve* translation quality
Comparing Hypotheses

• Comparing hypotheses with *same number of foreign words* covered

  Maria no dio una bofetada a la bruja verde

  e: Mary did not
  f: **------
  p: 0.154

  better
  partial
  translation

  e: the
  f: ------**--
  p: 0.354

  covers
  easier part
  --> lower cost

• Hypothesis that covers *easy part* of sentence is preferred

  Need to consider *future cost* of uncovered parts

  or: have one hypothesis stack per coverage vector
Synchronous Grammars

- Just like monolingual grammars except...
  - Each rule involves pairs (tuples) of nonterminals
  - Tuples of elementary trees for TAG, etc.
- First proposed for source-source translation in compilers
- Can be constituency, dependency, lexicalized, etc.
- Parsing speedups for monolingual grammar don’t necessarily work
  - E.g., no split-head trick for lexicalized parsing
- Binarization less straightforward
Bilingual Parsing

A variant of CKY chart parsing.

<table>
<thead>
<tr>
<th></th>
<th>póll’</th>
<th>oîd’</th>
<th>alôpēx</th>
</tr>
</thead>
<tbody>
<tr>
<td>the</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>fox</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>knows</td>
<td></td>
<td></td>
<td>NN/NN</td>
</tr>
<tr>
<td>many</td>
<td></td>
<td>VB/VB</td>
<td></td>
</tr>
<tr>
<td>things</td>
<td></td>
<td>JJ/JJ</td>
<td></td>
</tr>
</tbody>
</table>
Bilingual Parsing

<table>
<thead>
<tr>
<th></th>
<th>póll'</th>
<th>oîd'</th>
<th>alôpêx</th>
</tr>
</thead>
<tbody>
<tr>
<td>the</td>
<td></td>
<td></td>
<td>NP/NP</td>
</tr>
<tr>
<td>fox</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>knows</td>
<td></td>
<td>VP/VP</td>
<td></td>
</tr>
<tr>
<td>many</td>
<td></td>
<td>NP/NP</td>
<td></td>
</tr>
<tr>
<td>things</td>
<td></td>
<td>NP/NP</td>
<td></td>
</tr>
</tbody>
</table>

Diagram:

```
the ———— fox ———— knows
NP    NP    NP

NP    V'    NP
póll' oîd' alôpêx

NP    V'    NP
things
```
Bilingual Parsing

The diagram shows a sentence structure with the following components:

- The sentence is "The fox knows many things." in English.
- The sentence is translated into " póll’ oîd’ alôpêx" in a different language.

The table on the right side of the diagram maps the English sentence to its corresponding parts in the other language:

- "The" maps to " póll’".
- "Fox" maps to " oîd’".
- "Knows" maps to " alôpêx".
- "Many" maps to " the".
- "Things" maps to " many".

This diagram illustrates how bilingual parsing can be used to translate sentences from one language to another, preserving the grammatical structure of the original sentence.
Bilingual Parsing

The fox knows many things.

<table>
<thead>
<tr>
<th>póll’</th>
<th>oîd’</th>
<th>alôpêx</th>
</tr>
</thead>
<tbody>
<tr>
<td>the</td>
<td></td>
<td></td>
</tr>
<tr>
<td>fox</td>
<td></td>
<td></td>
</tr>
<tr>
<td>knows</td>
<td></td>
<td></td>
</tr>
<tr>
<td>many</td>
<td></td>
<td></td>
</tr>
<tr>
<td>things</td>
<td></td>
<td>S/S</td>
</tr>
</tbody>
</table>
MT as Parsing

• If we only have the source, parse it while recording all compatible target language trees.
• Runtime is also multiplied by a grammar constant: one string could be a noun and a verb phrase
• Continuing problem of multiple hidden configurations (trees, instead of phrases) for one translation.
What Makes Search Hard?

• What we really want: the best (highest-scoring) translation
• What we get: the best translation/phrase segmentation/alignment
  – Even summing over all ways of segmenting one translation is hard.
• Most common approaches:
  – Ignore problem
  – Sum over top $j$ translation/segmentation/alignment triples to get top $k<<j$ translations
Redundancy in $n$-best Lists

Source: Da ich wenig Zeit habe, gehe ich sofort in medias res.