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.

$10! = 3{,}626{,}800$ $20! \approx 2.43 \times 10^{18}$ $30! \approx 2.65 \times 10^{32}$

- 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!

Search in Phrase Models

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

Translate in target language order to ease language modeling.

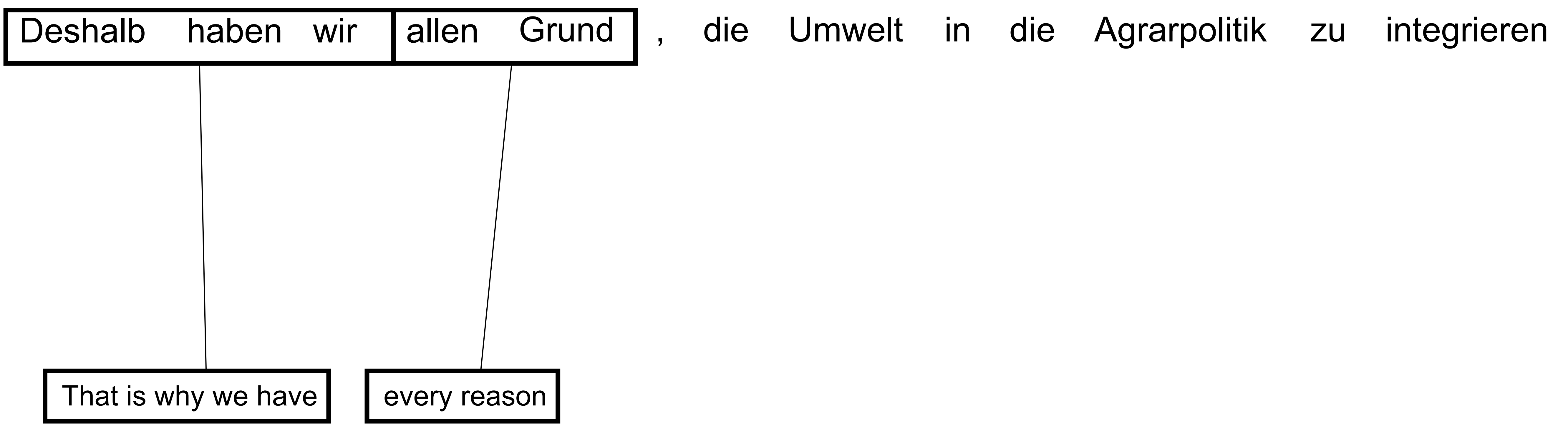
Search in Phrase Models

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

That is why we have

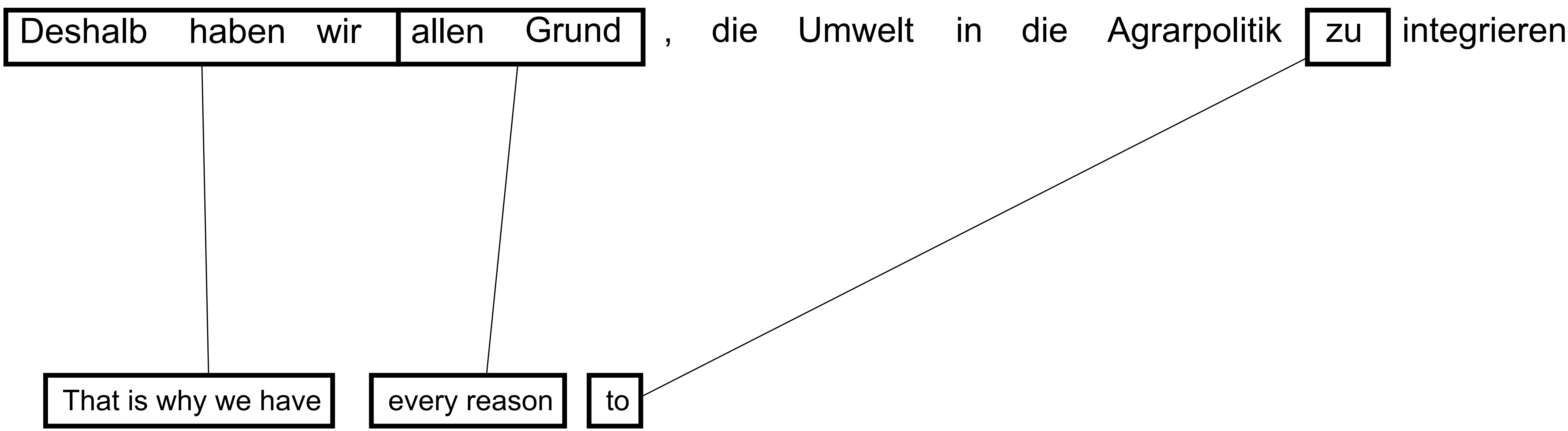
Translate in target language order to ease language modeling.

Search in Phrase Models



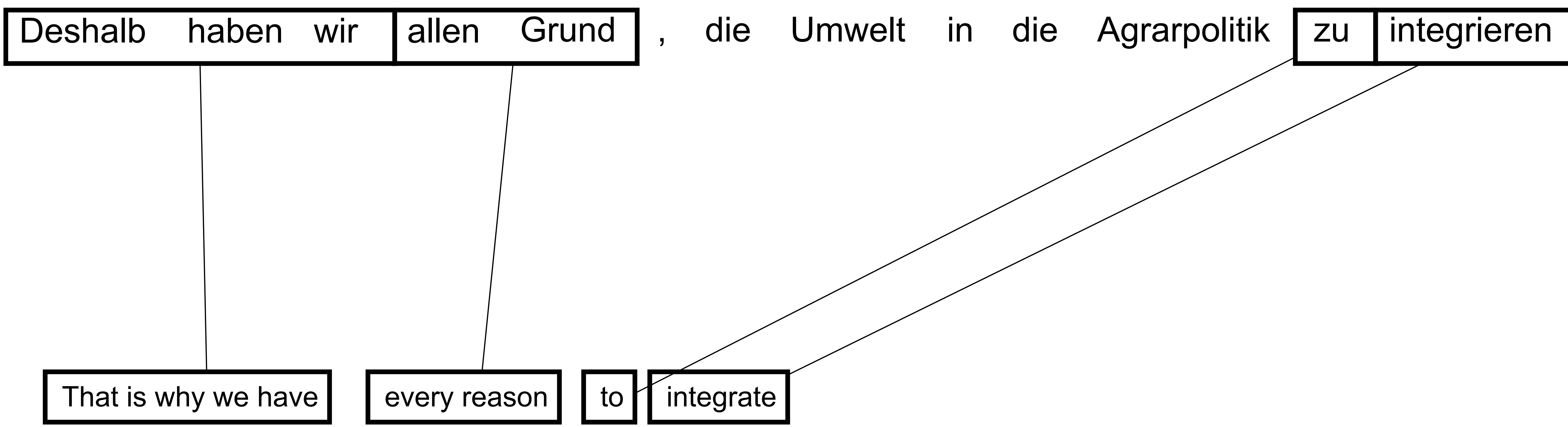
Translate in target language order to ease language modeling.

Search in Phrase Models



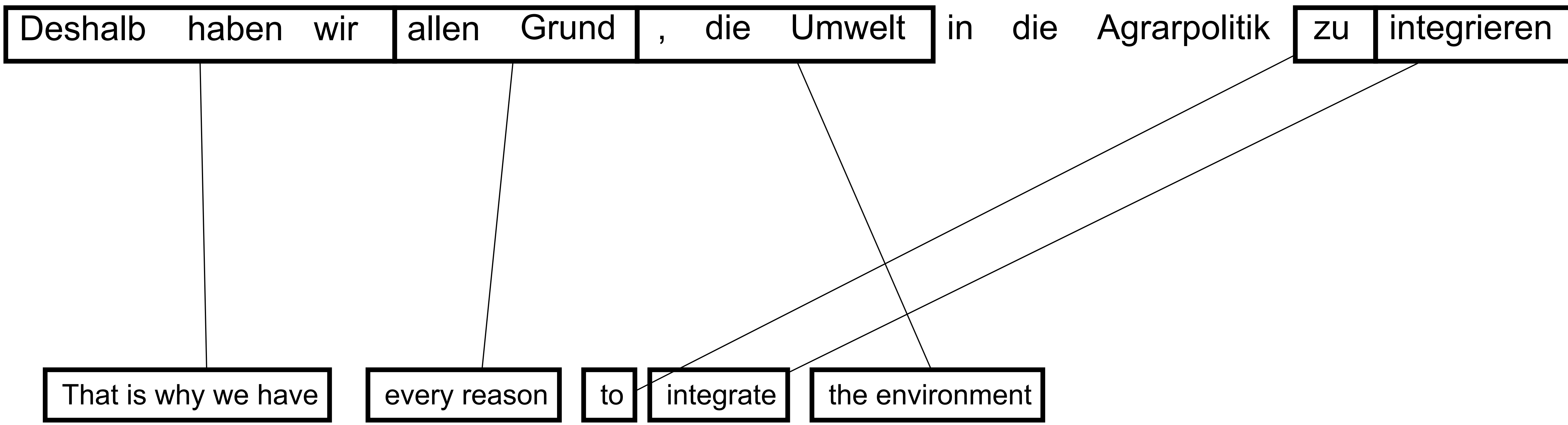
Translate in target language order to ease language modeling.

Search in Phrase Models



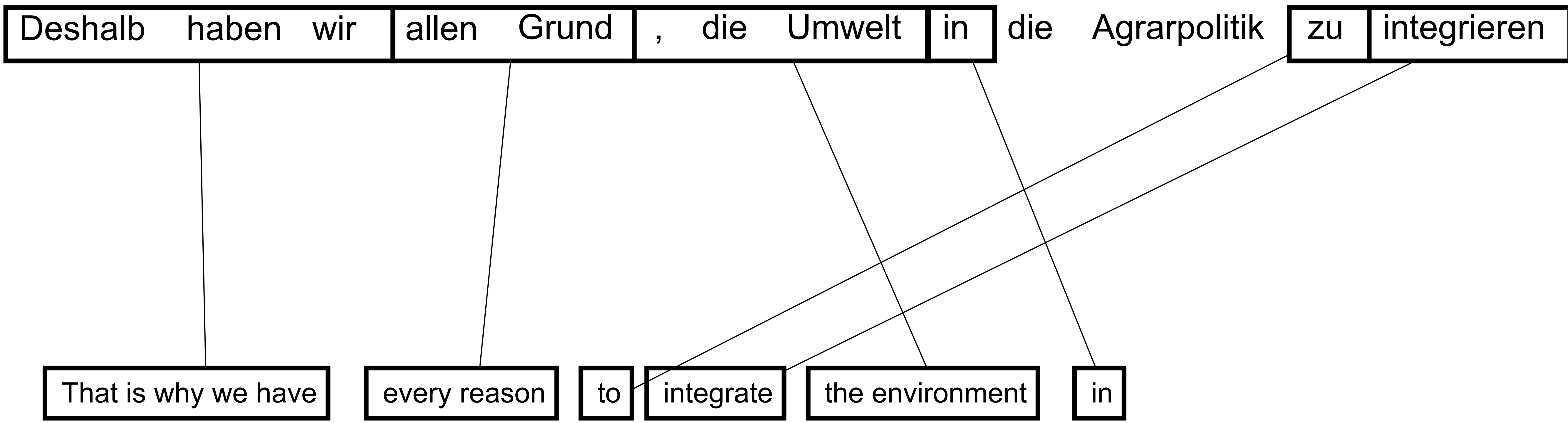
Translate in target language order to ease language modeling.

Search in Phrase Models



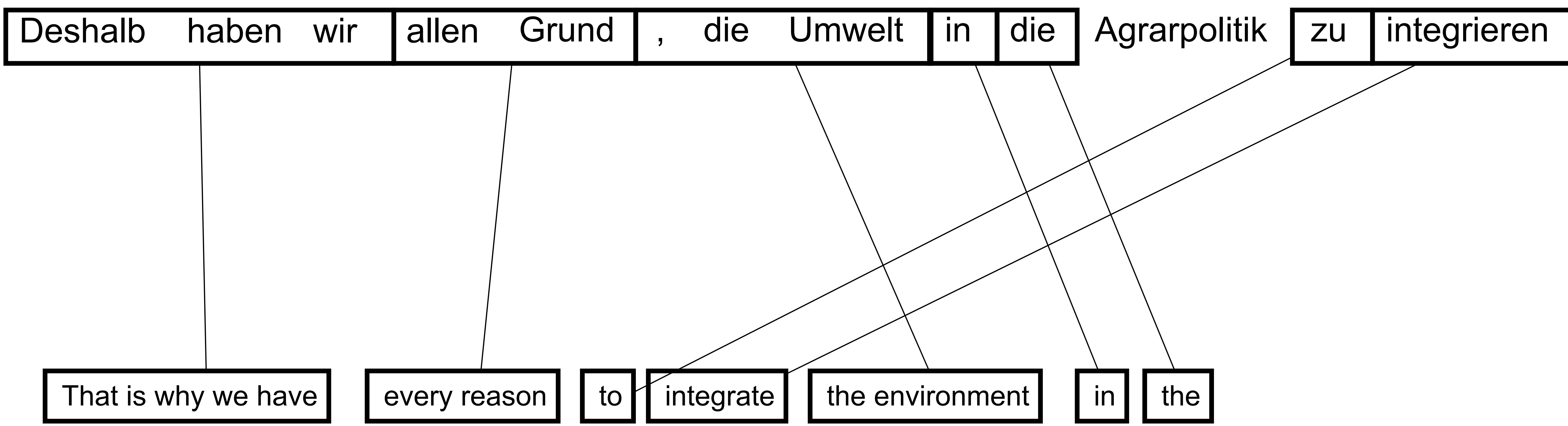
Translate in target language order to ease language modeling.

Search in Phrase Models



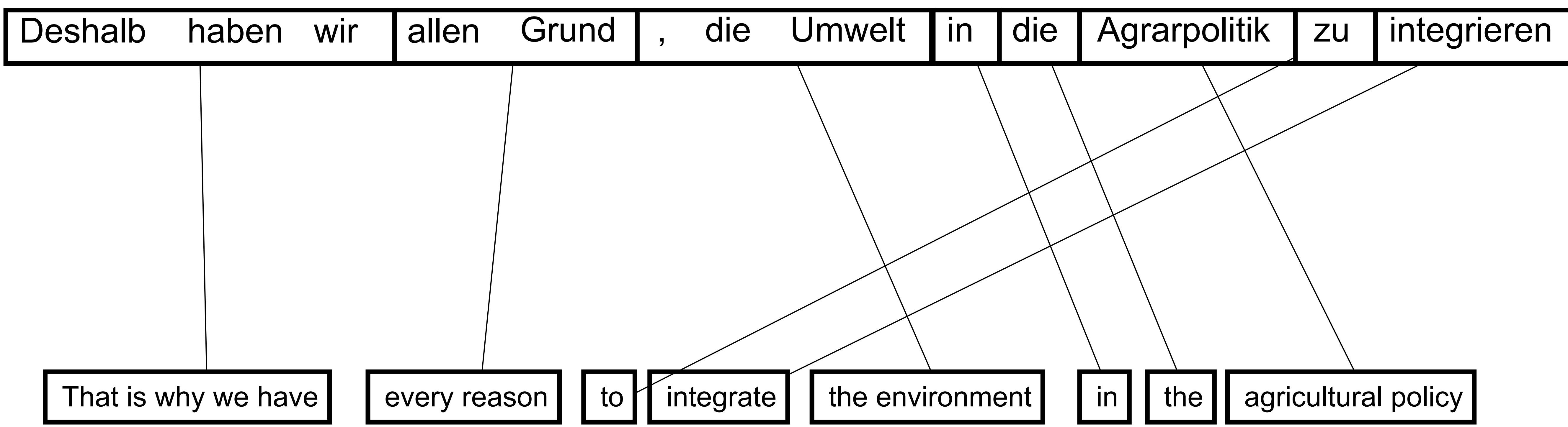
Translate in target language order to ease language modeling.

Search in Phrase Models



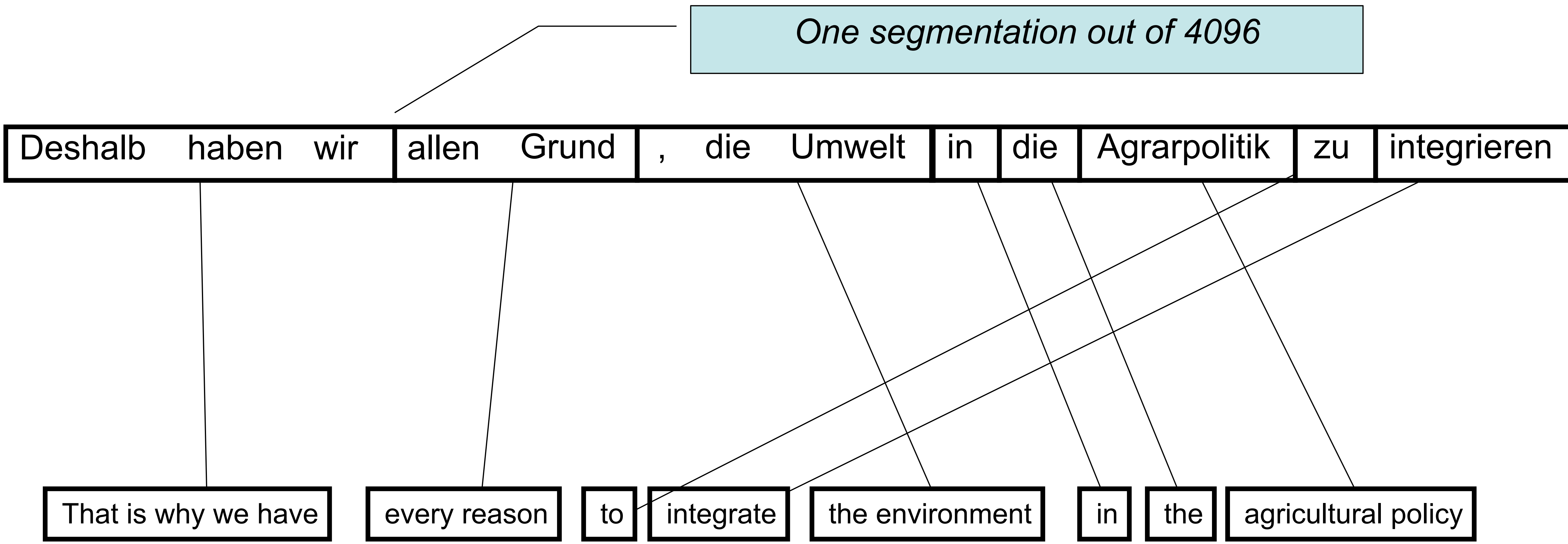
Translate in target language order to ease language modeling.

Search in Phrase Models



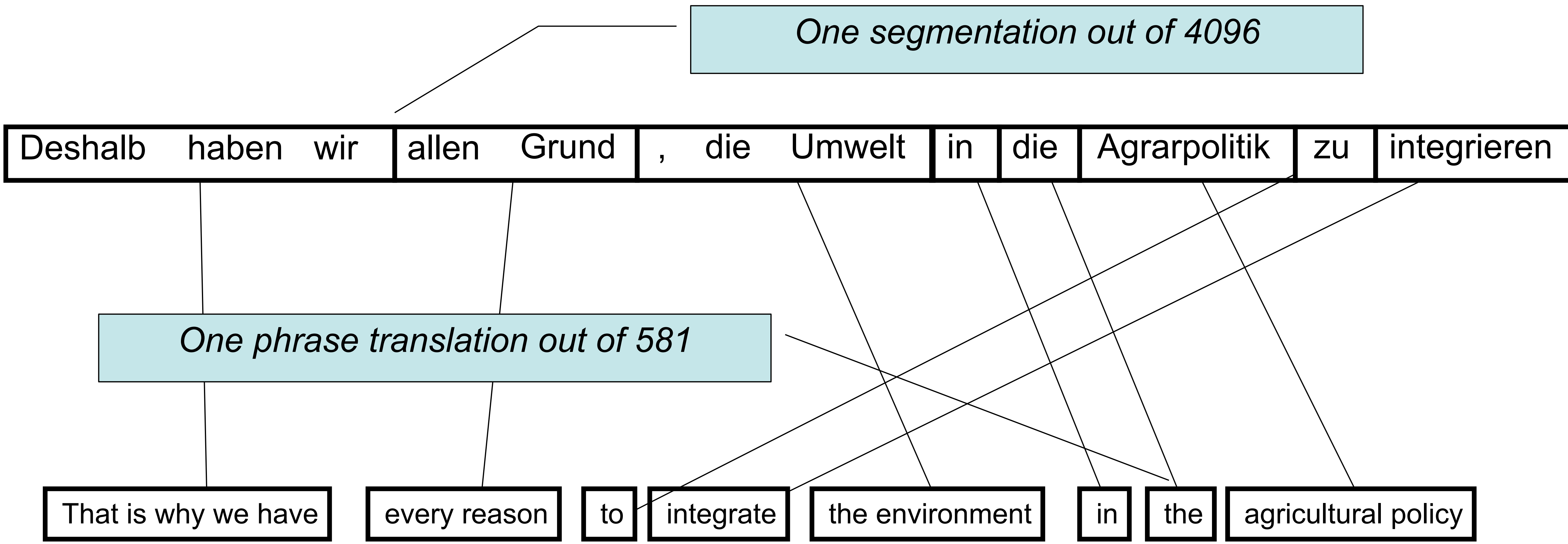
Translate in target language order to ease language modeling.

Search in Phrase Models



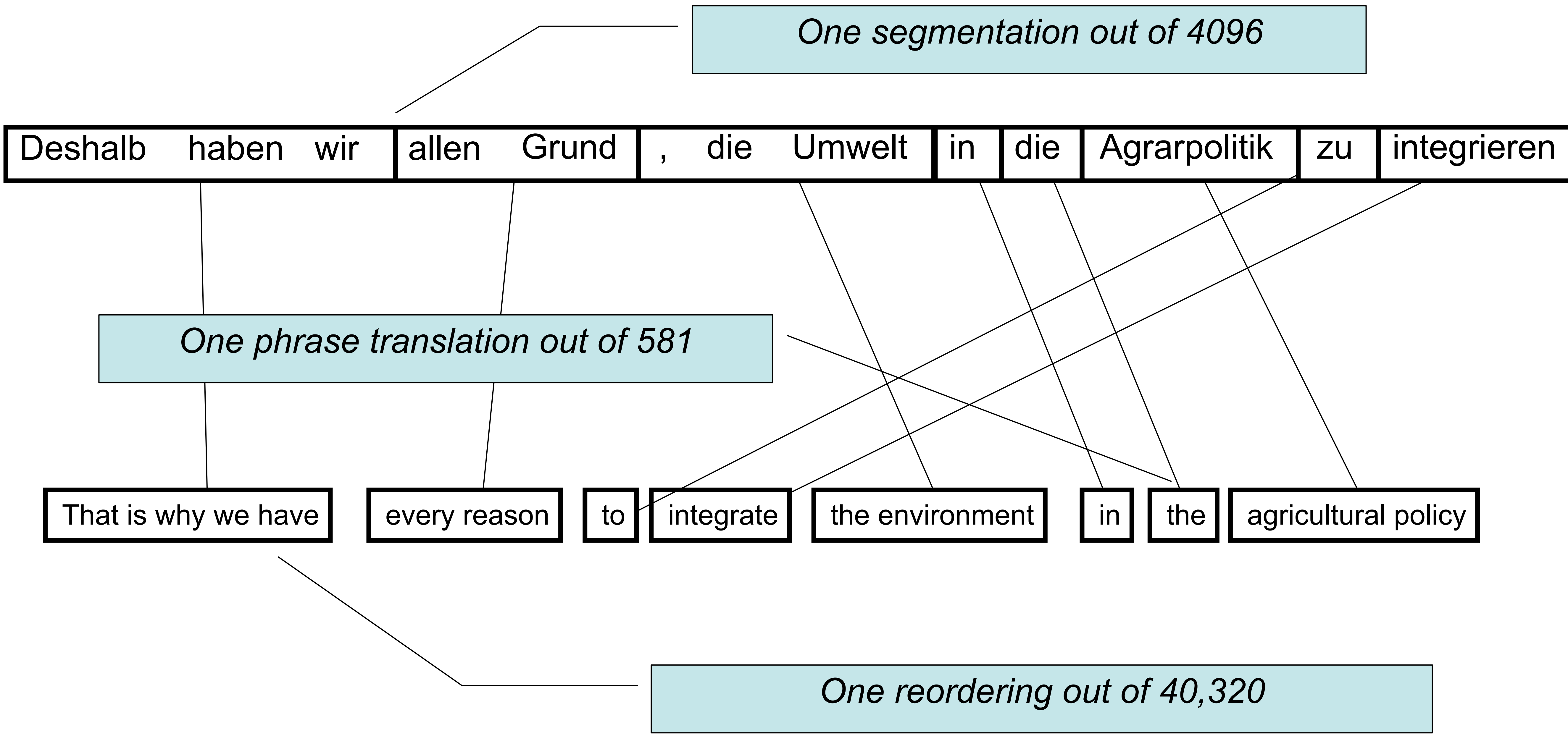
Translate in target language order to ease language modeling.

Search in Phrase Models



Translate in target language order to ease language modeling.

Search in Phrase Models



Translate in target language order to ease language modeling.

Search in Phrase Models

<table>
<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>
<tr><td colspan="3">that is why we have</td><td colspan="2">every reason</td><td colspan="3">the environment</td><td>in</td><td>the</td><td>agricultural policy</td><td>to</td><td>integrate</td></tr>
<tr><td>therefore</td><td>have</td><td>we</td><td colspan="3">every reason</td><td>the</td><td>environment</td><td colspan="2">in the</td><td>agricultural policy ,</td><td colspan="2">to integrate</td></tr>
<tr><td>that is why</td><td colspan="2">we have</td><td>all</td><td>reason</td><td>,</td><td>which</td><td colspan="2">environment in</td><td colspan="2">agricultural policy</td><td colspan="2">parliament</td></tr>
<tr><td colspan="2">have therefore</td><td>us</td><td>all the</td><td colspan="2">reason</td><td>of the</td><td colspan="2">environment into</td><td colspan="2">the agricultural policy</td><td colspan="2">successfully integrated</td></tr>
<tr><td colspan="2">hence</td><td>, we</td><td>every</td><td colspan="3">reason to make</td><td>environmental</td><td>on</td><td colspan="2">the cap</td><td colspan="2">be woven together</td></tr>
<tr><td colspan="3">we have therefore</td><td>everyone</td><td colspan="3">grounds for taking the</td><td>the environment</td><td colspan="2">to the</td><td>agricultural policy is</td><td>on</td><td>parliament</td></tr>
<tr><td>so</td><td colspan="2">, we</td><td>all of</td><td>cause</td><td colspan="2">which</td><td>environment ,</td><td>to</td><td colspan="2">the cap ,</td><td>for</td><td>incorporated</td></tr>
<tr><td colspan="3">hence our</td><td>any</td><td>why</td><td colspan="2">that</td><td>outside</td><td>at</td><td colspan="2">agricultural policy</td><td>too</td><td>woven together</td></tr>
<tr><td colspan="2">therefore ,</td><td>it</td><td>of all</td><td>reason for</td><td colspan="2">, the</td><td>completion</td><td>into</td><td colspan="2">that agricultural policy</td><td>be</td><td></td></tr>
</table>

And many, many more…even before reordering

Search in Phrase Models

<table>
<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>
<tr><td colspan="3">that is why we have</td><td colspan="2">every reason</td><td colspan="3">the environment</td><td>in</td><td>the</td><td>agricultural policy</td><td>to</td><td>integrate</td></tr>
<tr><td>therefore</td><td>have</td><td>we</td><td colspan="3">every reason</td><td>the</td><td>environment</td><td colspan="2">in the</td><td>agricultural policy ,</td><td colspan="2">to integrate</td></tr>
<tr><td>that is why</td><td colspan="2">we have</td><td>all</td><td>reason</td><td>,</td><td>which</td><td colspan="2">environment in</td><td colspan="2">agricultural policy</td><td colspan="2">parliament</td></tr>
<tr><td colspan="2">have therefore</td><td>us</td><td>all the</td><td colspan="2">reason</td><td>of the</td><td colspan="2">environment into</td><td colspan="2">the agricultural policy</td><td colspan="2">successfully integrated</td></tr>
<tr><td colspan="2">hence</td><td>, we</td><td>every</td><td colspan="3">reason to make</td><td>environmental</td><td>on</td><td colspan="2">the cap</td><td colspan="2">be woven together</td></tr>
<tr><td colspan="3">we have therefore</td><td>everyone</td><td colspan="3">grounds for taking the</td><td>the environment</td><td colspan="2">to the</td><td>agricultural policy is</td><td>on</td><td>parliament</td></tr>
<tr><td>so</td><td colspan="2">, we</td><td>all of</td><td>cause</td><td colspan="2">which</td><td>environment ,</td><td>to</td><td colspan="2">the cap ,</td><td>for</td><td>incorporated</td></tr>
<tr><td colspan="3">hence our</td><td>any</td><td>why</td><td colspan="2">that</td><td>outside</td><td>at</td><td colspan="2">agricultural policy</td><td>too</td><td>woven together</td></tr>
<tr><td colspan="2">therefore ,</td><td>it</td><td>of all</td><td>reason for</td><td colspan="2">, the</td><td>completion</td><td>into</td><td colspan="2">that agricultural policy</td><td>be</td><td></td></tr>
</table>

And many, many more…even before reordering

Search in Phrase Models

<table>
<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>
<tr><td colspan="3">that is why we have</td><td colspan="2">every reason</td><td colspan="3">the environment</td><td>in</td><td>the</td><td>agricultural policy</td><td>to</td><td>integrate</td></tr>
<tr><td>therefore</td><td>have</td><td>we</td><td colspan="3">every reason</td><td>the</td><td>environment</td><td colspan="2">in the</td><td>agricultural policy ,</td><td colspan="2">to integrate</td></tr>
<tr><td>that is why</td><td colspan="2">we have</td><td>all</td><td>reason</td><td>,</td><td>which</td><td colspan="2">environment in</td><td colspan="2">agricultural policy</td><td colspan="2">parliament</td></tr>
<tr><td colspan="2">have therefore</td><td>us</td><td>all the</td><td colspan="2">reason</td><td>of the</td><td colspan="2">environment into</td><td colspan="2">the agricultural policy</td><td colspan="2">successfully integrated</td></tr>
<tr><td colspan="2">hence</td><td>, we</td><td>every</td><td colspan="3">reason to make</td><td>environmental</td><td>on</td><td colspan="2">the cap</td><td colspan="2">be woven together</td></tr>
<tr><td colspan="3">we have therefore</td><td>everyone</td><td colspan="3">grounds for taking the</td><td>the environment</td><td colspan="2">to the</td><td>agricultural policy is</td><td>on</td><td>parliament</td></tr>
<tr><td>so</td><td colspan="2">, we</td><td>all of</td><td>cause</td><td colspan="2">which</td><td>environment ,</td><td>to</td><td colspan="2">the cap ,</td><td>for</td><td>incorporated</td></tr>
<tr><td colspan="3">hence our</td><td>any</td><td>why</td><td colspan="2">that</td><td>outside</td><td>at</td><td colspan="2">agricultural policy</td><td>too</td><td>woven together</td></tr>
<tr><td colspan="2">therefore ,</td><td>it</td><td>of all</td><td>reason for</td><td colspan="2">, the</td><td>completion</td><td>into</td><td colspan="2">that agricultural policy</td><td>be</td><td></td></tr>
</table>

And many, many more…even before reordering

Search in Phrase Models

<table>
<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>
<tr><td colspan="3">that is why we have</td><td colspan="2">every reason</td><td colspan="3">the environment</td><td>in</td><td>the</td><td>agricultural policy</td><td>to</td><td>integrate</td></tr>
<tr><td>therefore</td><td>have</td><td>we</td><td colspan="3">every reason</td><td>the</td><td>environment</td><td colspan="2">in the</td><td>agricultural policy ,</td><td colspan="2">to integrate</td></tr>
<tr><td>that is why</td><td colspan="2">we have</td><td>all</td><td>reason</td><td>,</td><td>which</td><td colspan="2">environment in</td><td colspan="2">agricultural policy</td><td colspan="2">parliament</td></tr>
<tr><td colspan="2">have therefore</td><td>us</td><td>all the</td><td colspan="2">reason</td><td>of the</td><td colspan="2">environment into</td><td colspan="2">the agricultural policy</td><td colspan="2">successfully integrated</td></tr>
<tr><td colspan="2">hence</td><td>, we</td><td>every</td><td colspan="3">reason to make</td><td>environmental</td><td>on</td><td colspan="2">the cap</td><td colspan="2">be woven together</td></tr>
<tr><td colspan="3">we have therefore</td><td>everyone</td><td colspan="3">grounds for taking the</td><td>the environment</td><td colspan="2">to the</td><td>agricultural policy is</td><td>on</td><td>parliament</td></tr>
<tr><td>so</td><td colspan="2">, we</td><td>all of</td><td>cause</td><td colspan="2">which</td><td>environment ,</td><td>to</td><td colspan="2">the cap ,</td><td>for</td><td>incorporated</td></tr>
<tr><td colspan="3">hence our</td><td>any</td><td>why</td><td colspan="2">that</td><td>outside</td><td>at</td><td colspan="2">agricultural policy</td><td>too</td><td>woven together</td></tr>
<tr><td colspan="2">therefore ,</td><td>it</td><td>of all</td><td>reason for</td><td colspan="2">, the</td><td>completion</td><td>into</td><td colspan="2">that agricultural policy</td><td>be</td><td></td></tr>
</table>

And many, many more…even before reordering

Search in Phrase Models

<table>
<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>
<tr><td colspan="3">that is why we have</td><td colspan="2">every reason</td><td colspan="3">the environment</td><td>in</td><td>the</td><td>agricultural policy</td><td>to</td><td>integrate</td></tr>
<tr><td>therefore</td><td>have</td><td>we</td><td colspan="3">every reason</td><td>the</td><td>environment</td><td colspan="2">in the</td><td>agricultural policy ,</td><td colspan="2">to integrate</td></tr>
<tr><td>that is why</td><td colspan="2">we have</td><td>all</td><td>reason</td><td>,</td><td>which</td><td colspan="2">environment in</td><td colspan="2">agricultural policy</td><td colspan="2">parliament</td></tr>
<tr><td colspan="2">have therefore</td><td>us</td><td>all the</td><td colspan="2">reason</td><td>of the</td><td colspan="2">environment into</td><td colspan="2">the agricultural policy</td><td colspan="2">successfully integrated</td></tr>
<tr><td colspan="2">hence</td><td>, we</td><td>every</td><td colspan="3">reason to make</td><td>environmental</td><td>on</td><td colspan="2">the cap</td><td colspan="2">be woven together</td></tr>
<tr><td colspan="3">we have therefore</td><td>everyone</td><td colspan="3">grounds for taking the</td><td>the environment</td><td colspan="2">to the</td><td>agricultural policy is</td><td>on</td><td>parliament</td></tr>
<tr><td>so</td><td colspan="2">, we</td><td>all of</td><td>cause</td><td colspan="2">which</td><td>environment ,</td><td>to</td><td colspan="2">the cap ,</td><td>for</td><td>incorporated</td></tr>
<tr><td colspan="3">hence our</td><td>any</td><td>why</td><td colspan="2">that</td><td>outside</td><td>at</td><td colspan="2">agricultural policy</td><td>too</td><td>woven together</td></tr>
<tr><td colspan="2">therefore ,</td><td>it</td><td>of all</td><td>reason for</td><td colspan="2">, the</td><td>completion</td><td>into</td><td colspan="2">that agricultural policy</td><td>be</td><td></td></tr>
</table>

And many, many more…even before reordering

“Stack Decoding”

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

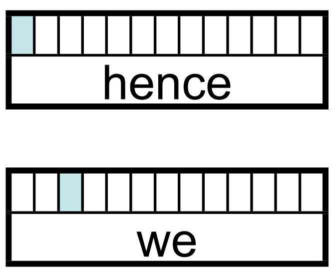
“Stack Decoding”

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

hence

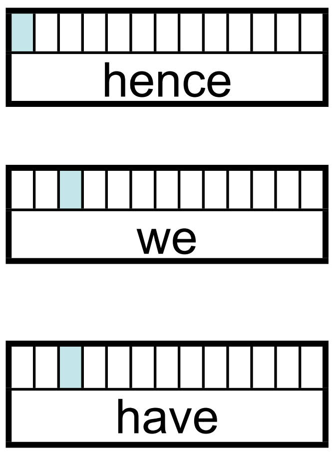
“Stack Decoding”

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



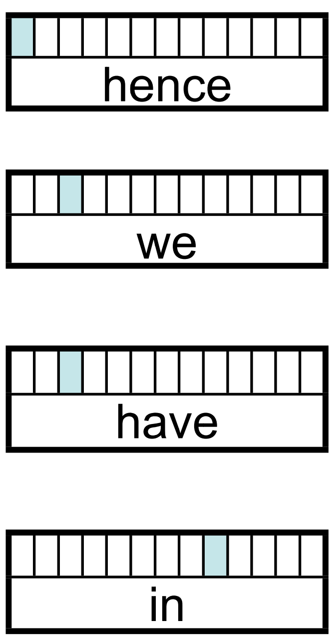
“Stack Decoding”

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



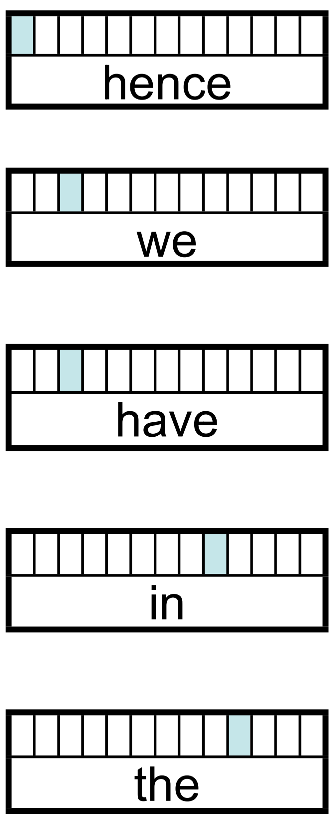
“Stack Decoding”

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



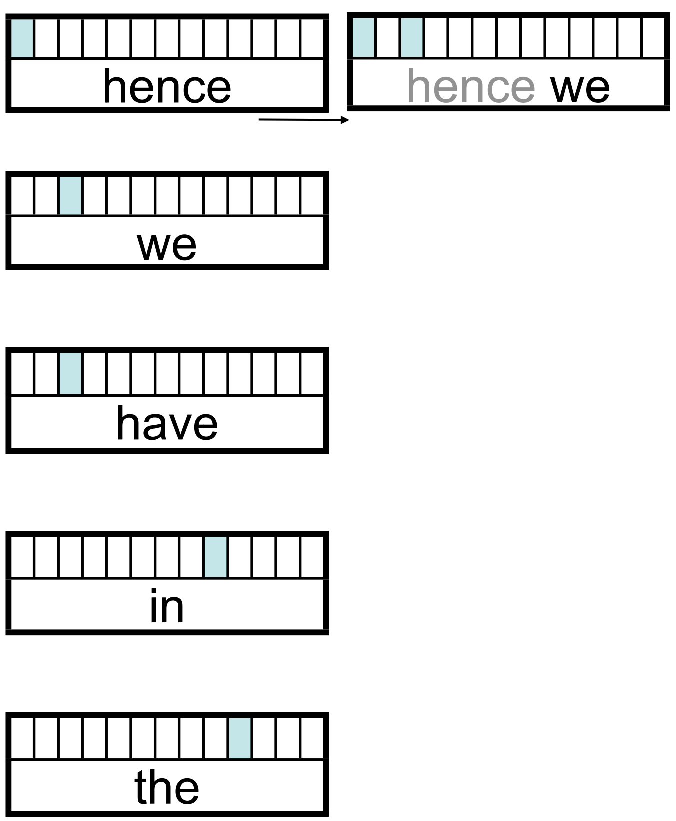
“Stack Decoding”

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



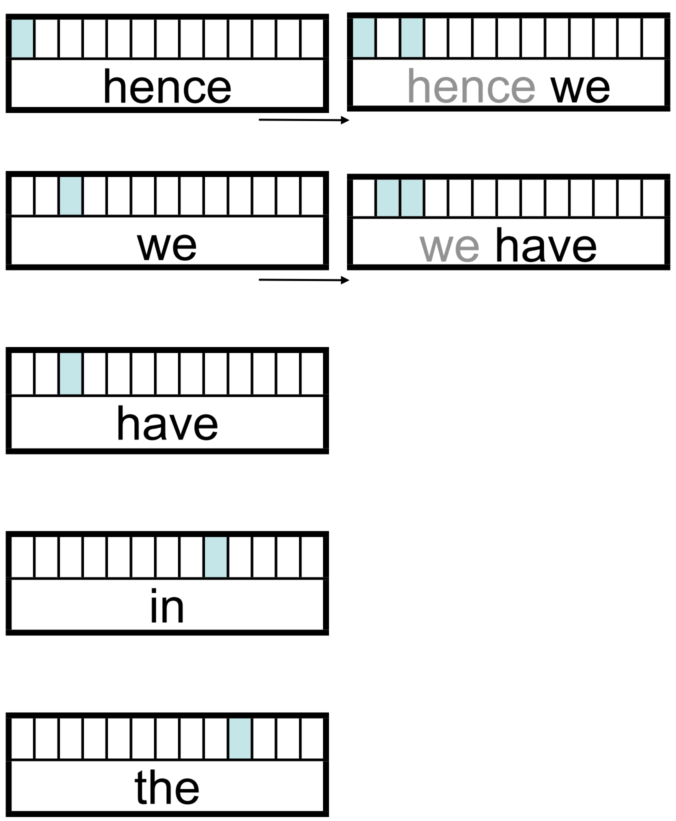
“Stack Decoding”

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



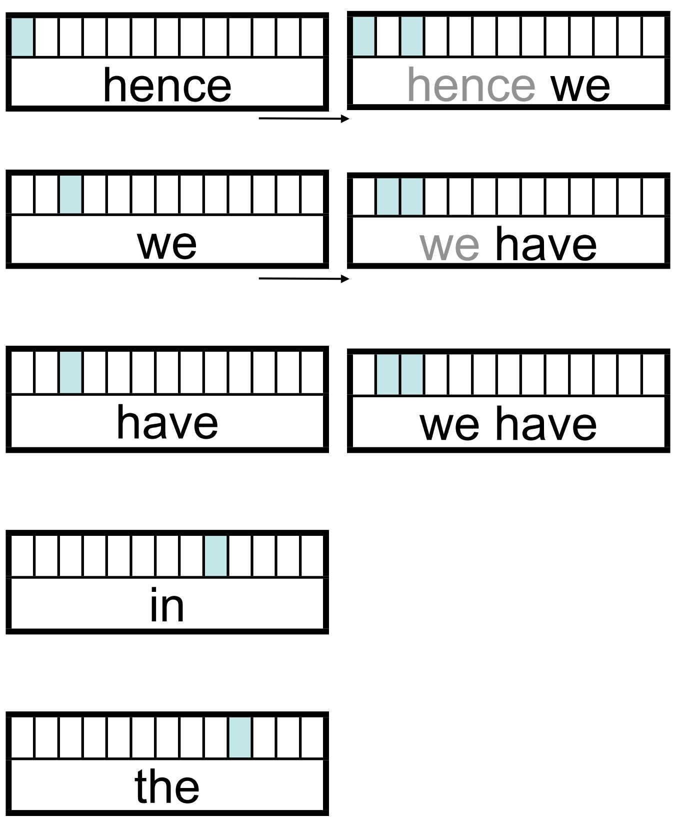
“Stack Decoding”

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



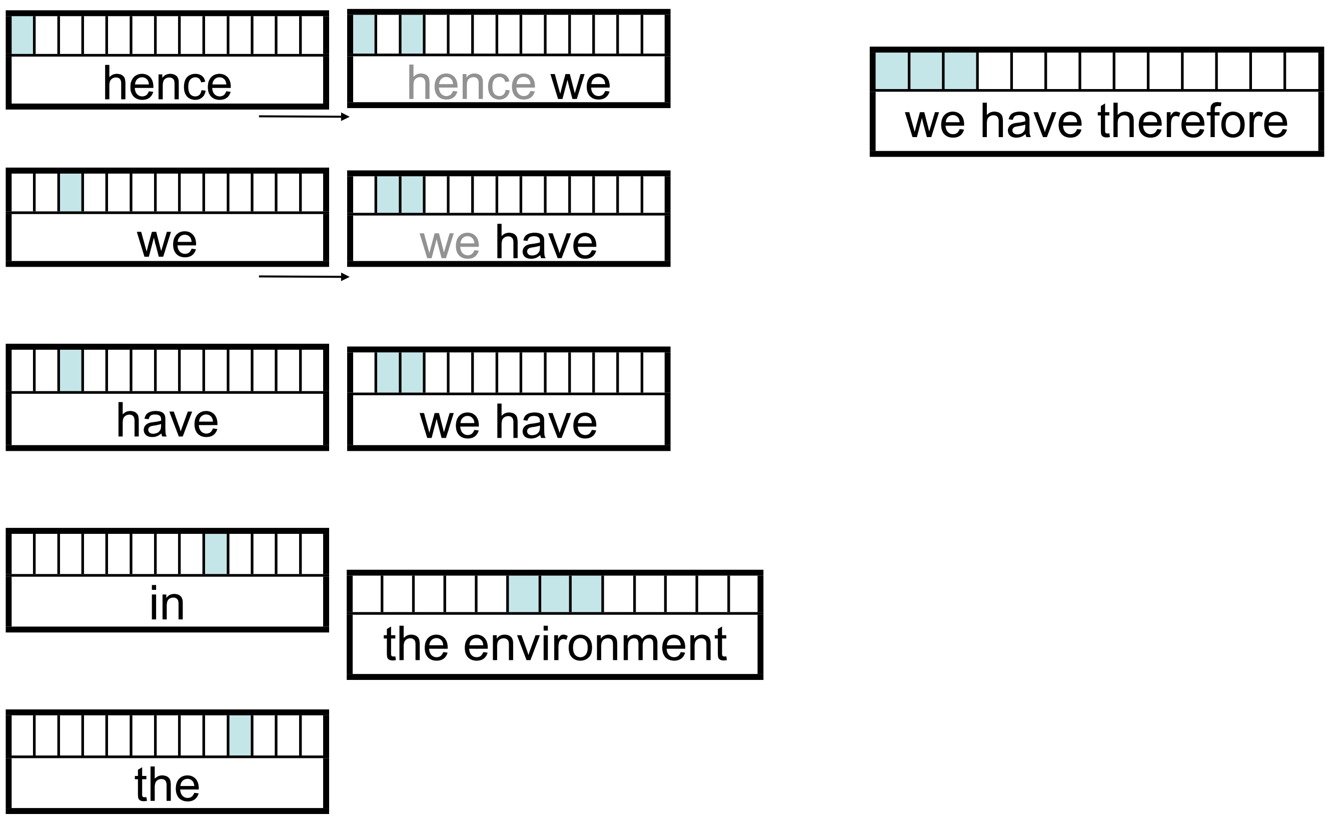
“Stack Decoding”

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



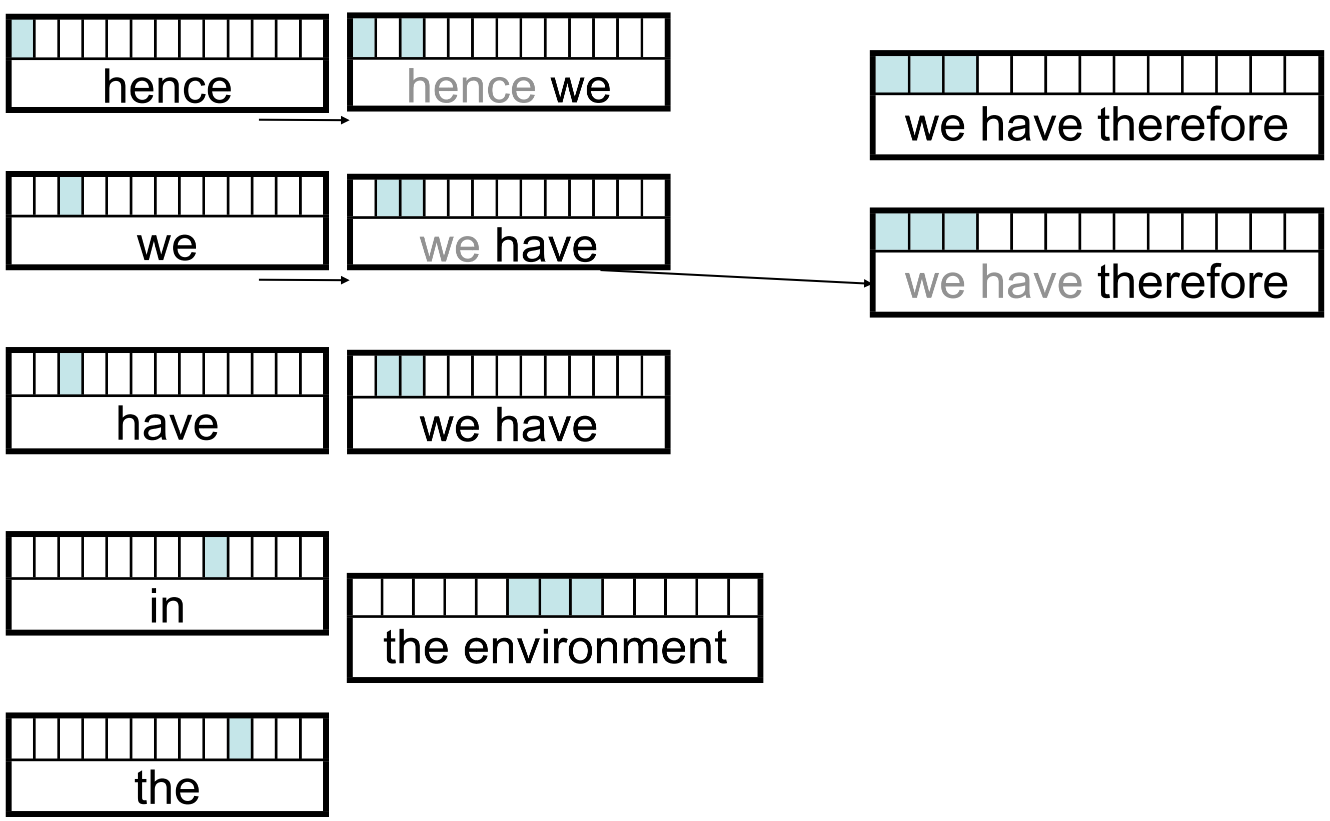
“Stack Decoding”

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



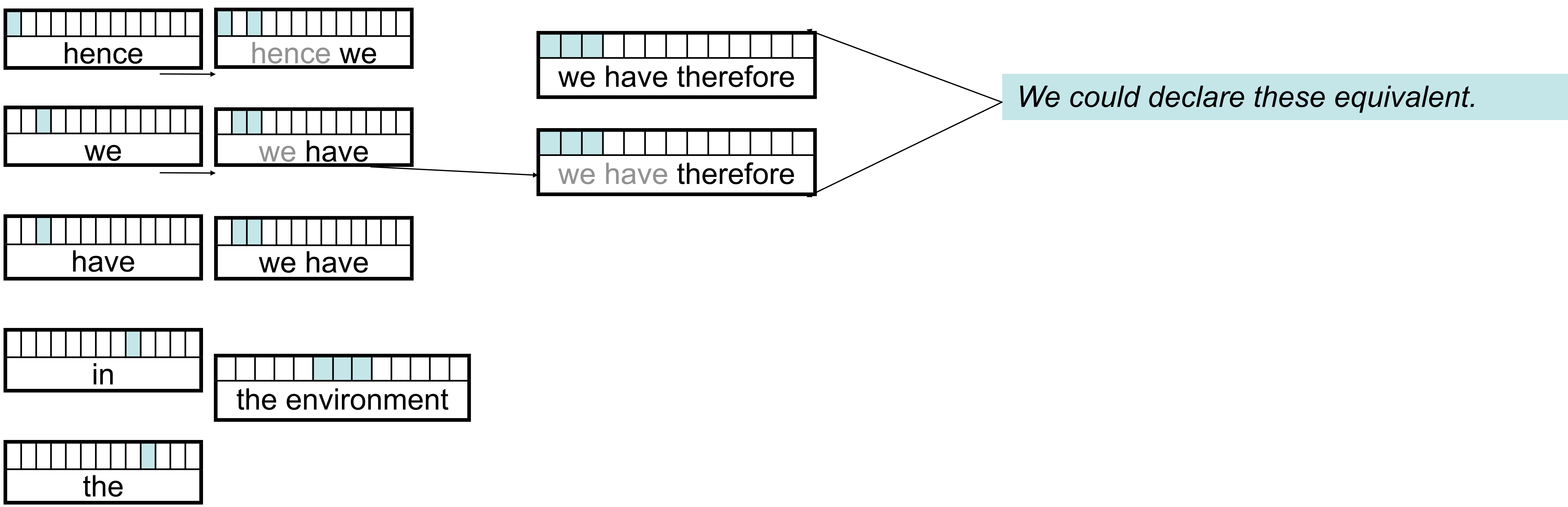
“Stack Decoding”

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



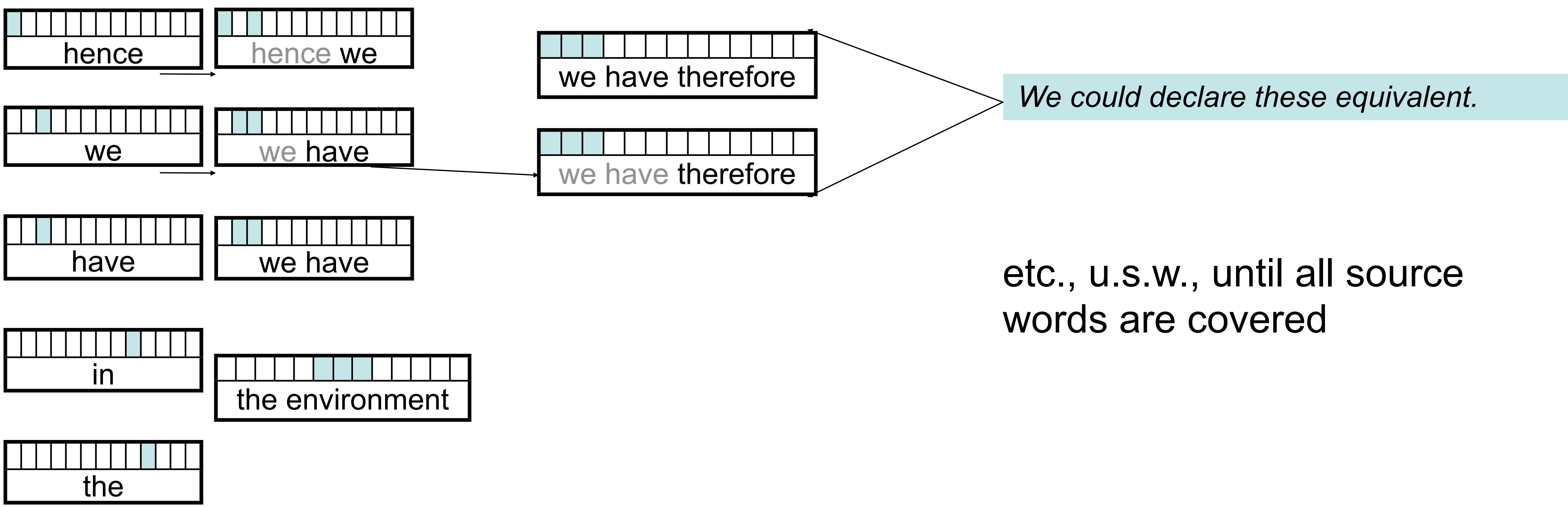
“Stack Decoding”

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



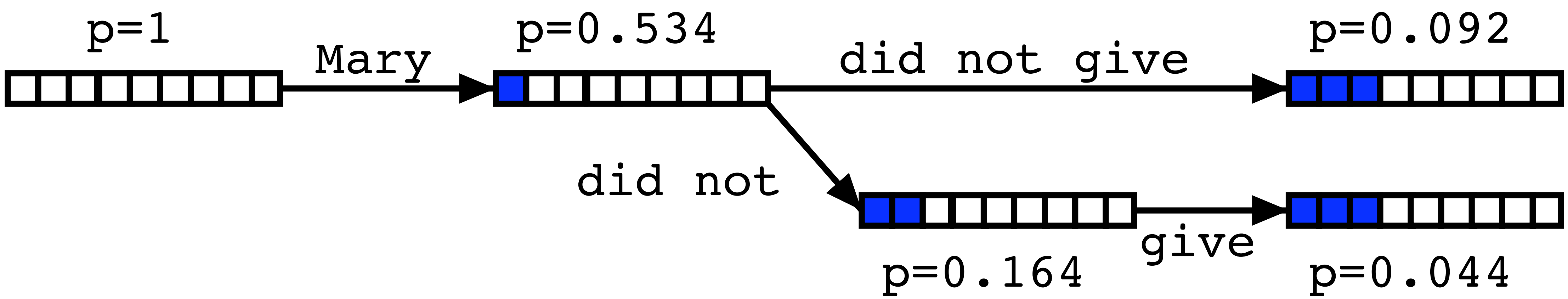
etc., u.s.w., until all source words are covered

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", …

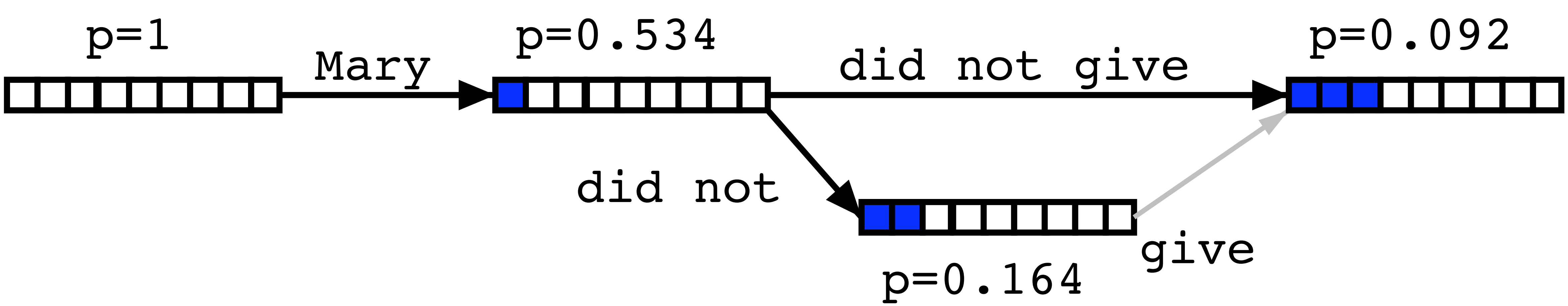
Hypothesis Recombination

- Different paths to the same partial translation



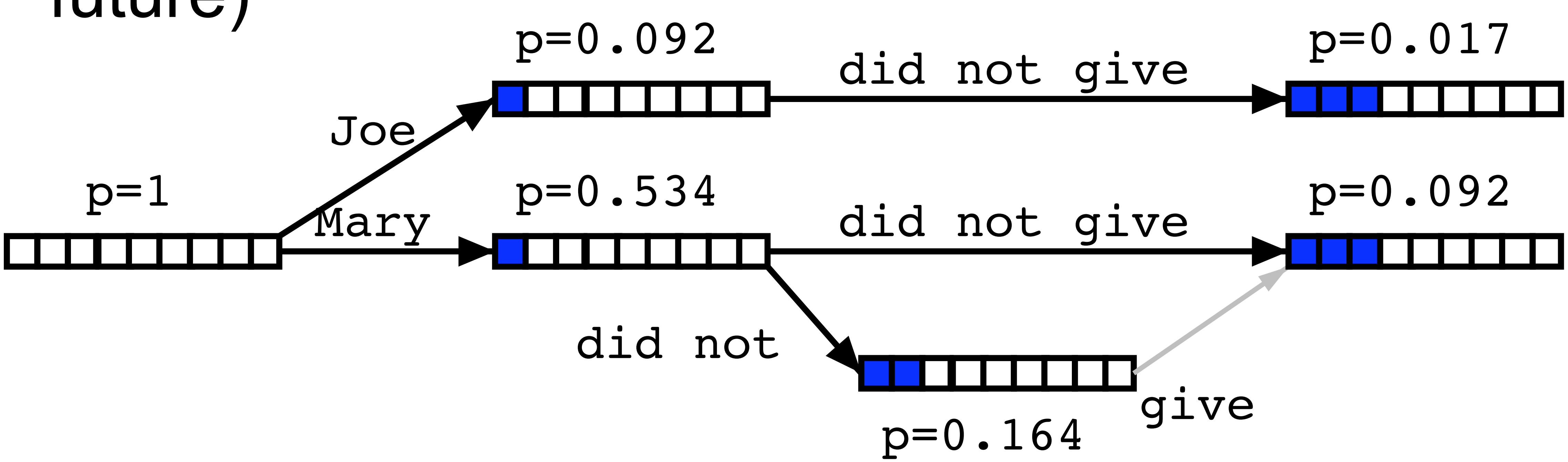
Hypothesis Recombination

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



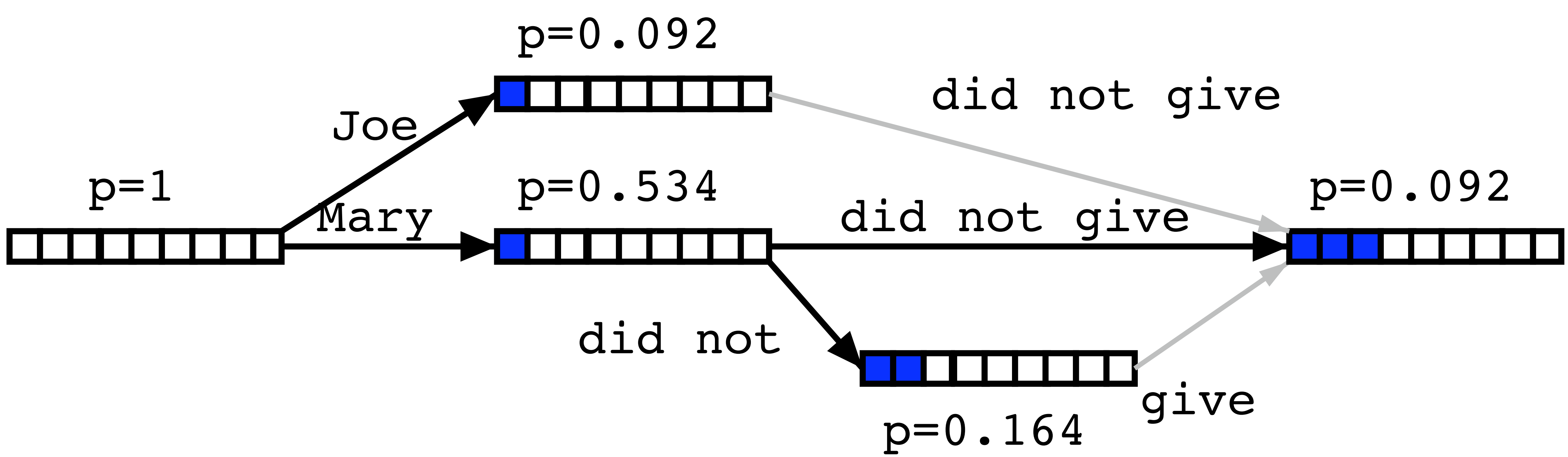
Hypothesis Recombination

- 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)



Hypothesis Recombination

- Combining partially matching hypotheses



Pruning

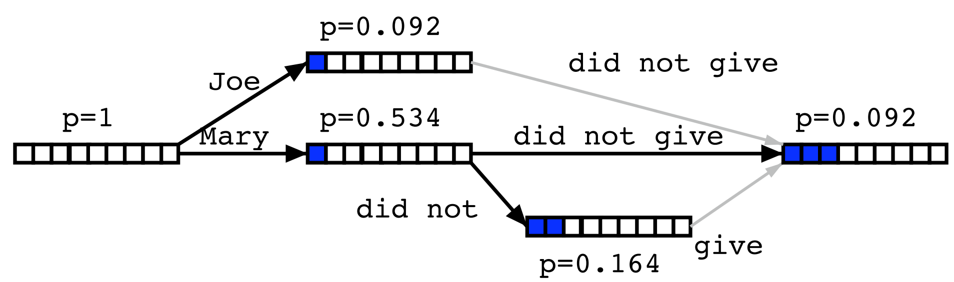
- Hypothesis recombination is *not su cient*

  Heuristically *discard* weak hypotheses early

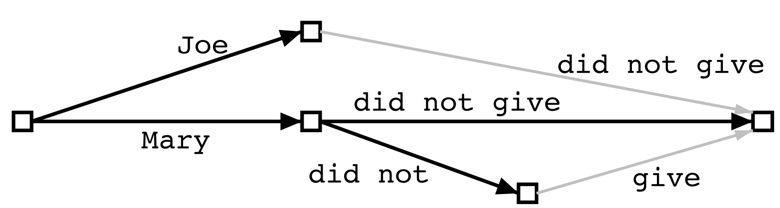
- Organize Hypothesis in **stacks**, e.g. by
  - *same* foreign words covered
  - *same number* of foreign words covered
  - *same number* of English words produced

- 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 times the cost of best hypothesis in stack (e.g., $= 0.001$)

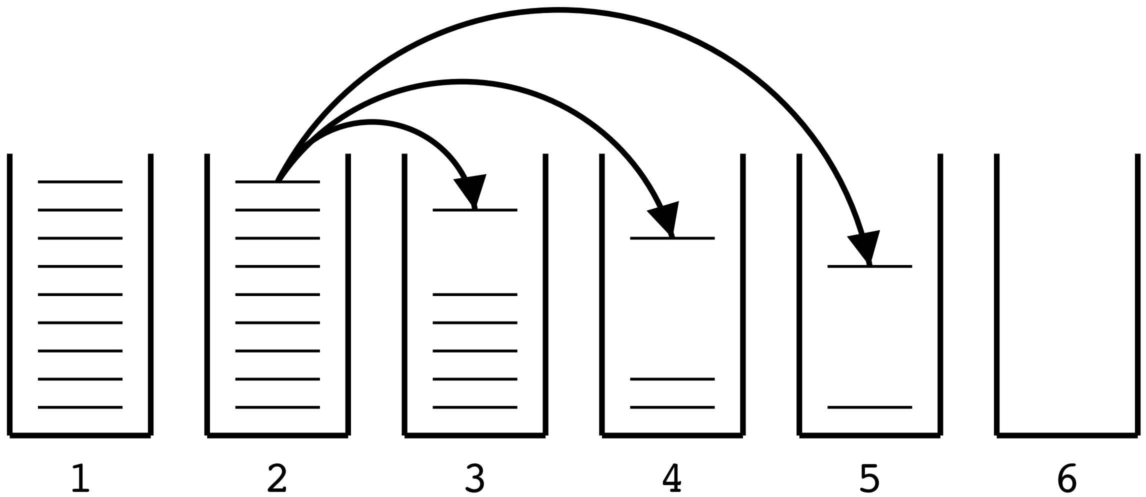
Word Lattice Generation



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



Hypothesis Stacks



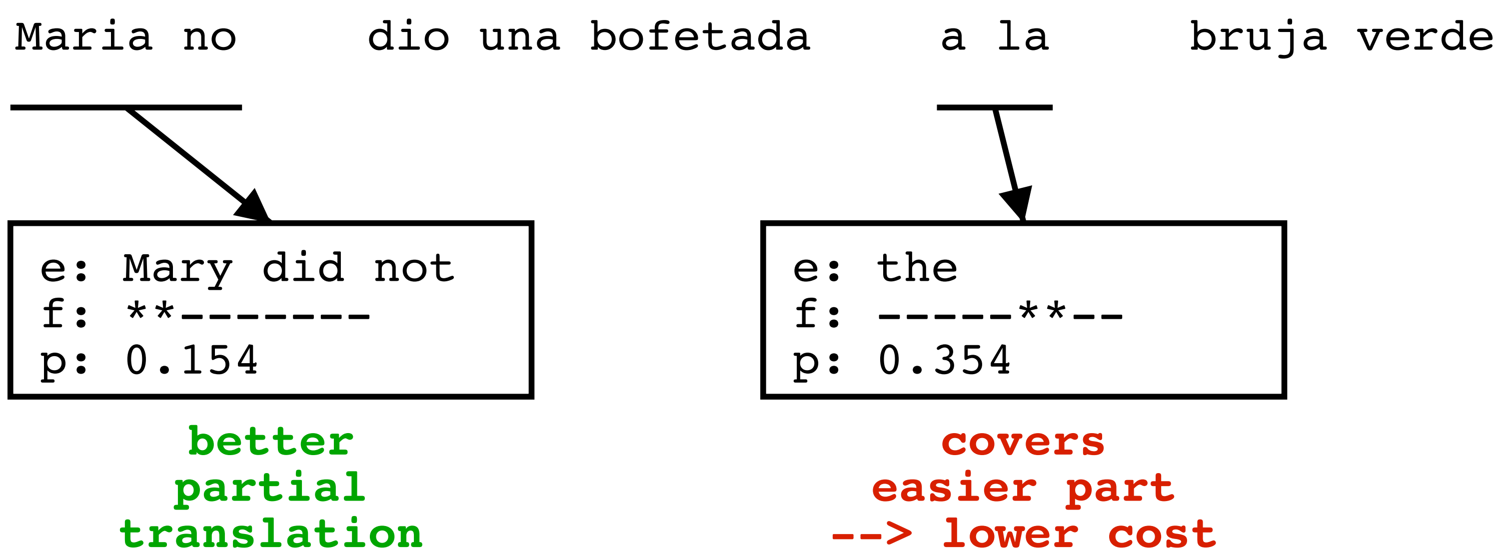
- Organization of hypothesis into stacks
  - here: based on *number of foreign words* translated
  - during translation all hypotheses from one stack are expanded
  - expanded Hypotheses are placed into stacks

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



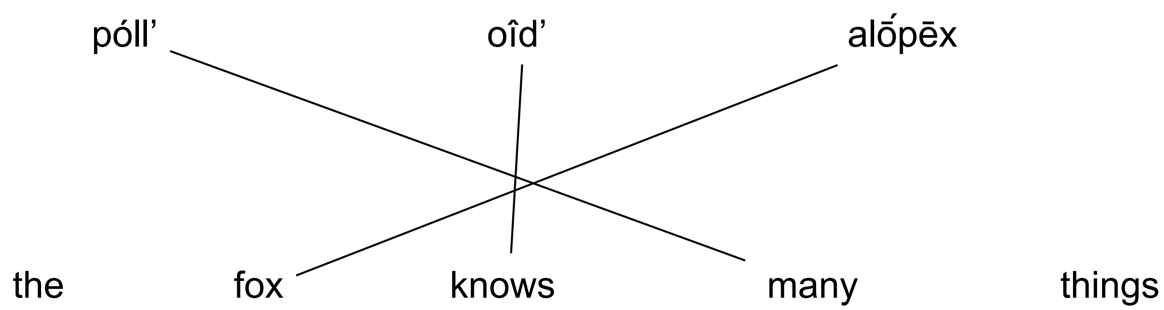
- Hypothesis that covers *easy part* of sentence is preferred

  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

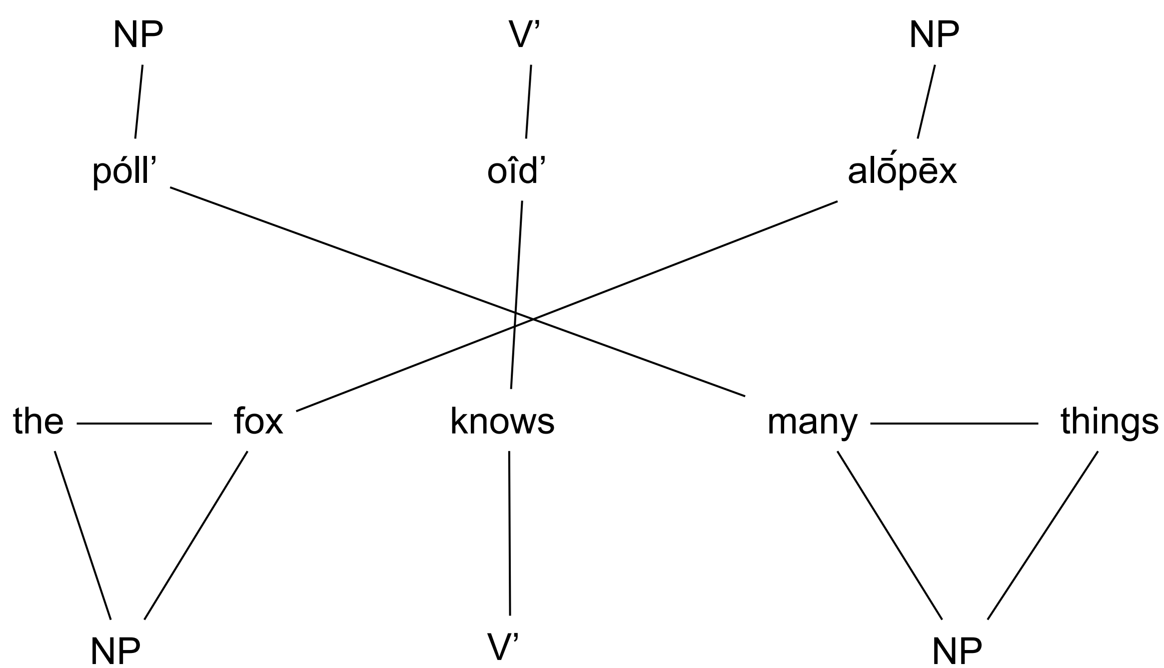
Bilingual Parsing



A variant of CKY chart parsing.

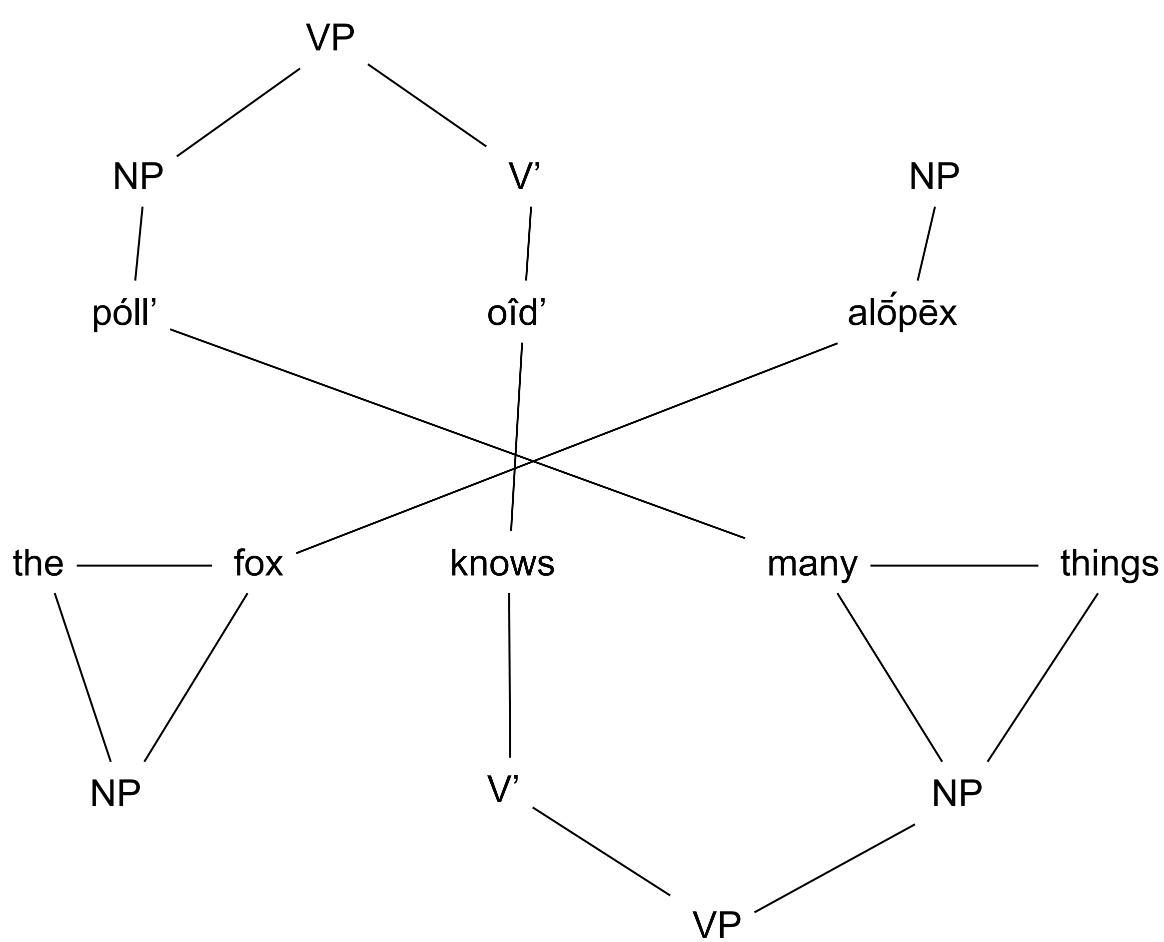
	póll’	oîd’	alṓpēx
the			
fox			NN/NN
knows		VB/VB	
many	JJ/JJ		
things			

Bilingual Parsing



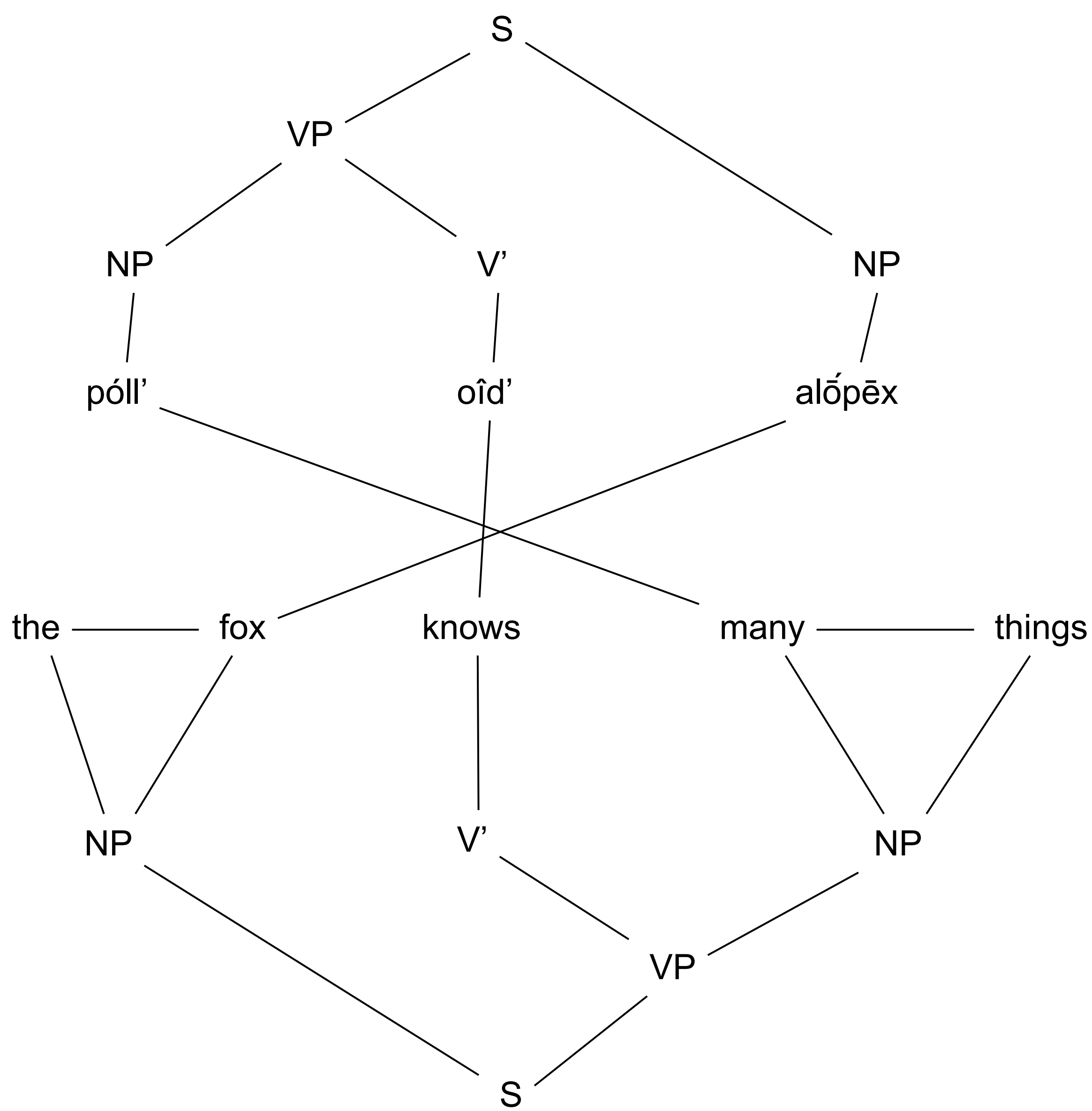
	póll’	oîd’	alṓpēx
the			NP/NP
fox			
knows		VP/VP	
many	NP/NP		
things			

Bilingual Parsing



<table>
<tr><th></th><th>póll’</th><th>oîd’</th><th>alṓpēx</th></tr>
<tr><td>the</td><td></td><td></td><td rowspan="2">NP/NP</td></tr>
<tr><td>fox</td><td></td><td></td></tr>
<tr><td>knows</td><td colspan="2" rowspan="3">VP/VP</td><td></td></tr>
<tr><td>many</td><td></td></tr>
<tr><td>things</td><td></td></tr>
</table>

Bilingual Parsing



	póll’	oîd’	alṓpēx
the			
fox			
knows		S/S	
many			
things			

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 .

as i have little time , i am immediately in medias res . | 0-1,0-1 2-2,4-4 3-4,2-3 5-5,5-5 6-7,6-7 8-8,8-8 9-9,9-9 10-10,10-10 11-11,11-11 12-12,12-12
as i have little time , i am immediately in medias res . | 0-0,0-0 1-1,1-1 2-2,4-4 3-4,2-3 5-5,5-5 6-7,6-7 8-8,8-8 9-9,9-9 10-10,10-10 11-11,11-11 12-12,12-12
as i have little time , i am in medias res immediately . | 0-1,0-1 2-2,4-4 3-4,2-3 5-5,5-5 6-7,6-7 8-8,9-9 9-9,10-10 10-10,11-11 11-11,8-8 12-12,12-12
as i have little time , i am in medias res immediately . | 0-0,0-0 1-1,1-1 2-2,4-4 3-4,2-3 5-5,5-5 6-7,6-7 8-8,9-9 9-9,10-10 10-10,11-11 11-11,8-8 12-12,12-12
as i have little time , i am immediately in medias res . | 0-1,0-1 2-2,4-4 3-3,2-2 4-4,3-3 5-5,5-5 6-7,6-7 8-8,8-8 9-9,9-9 10-10,10-10 11-11,11-11 12-12,12-12
as i have little time , i am immediately in medias res . | 0-0,0-0 1-1,1-1 2-2,4-4 3-3,2-2 4-4,3-3 5-5,5-5 6-7,6-7 8-8,8-8 9-9,9-9 10-10,10-10 11-11,11-11 12-12,12-12
as i have little time , i am in medias res immediately . | 0-1,0-1 2-2,4-4 3-3,2-2 4-4,3-3 5-5,5-5 6-7,6-7 8-8,9-9 9-9,10-10 10-10,11-11 11-11,8-8 12-12,12-12
as i have little time , i am in medias res immediately . | 0-0,0-0 1-1,1-1 2-2,4-4 3-3,2-2 4-4,3-3 5-5,5-5 6-7,6-7 8-8,9-9 9-9,10-10 10-10,11-11 11-11,8-8 12-12,12-12
as i have little time , i am immediately in medias res . | 0-1,0-1 2-2,4-4 3-4,2-3 5-5,5-5 6-6,7-7 7-7,6-6 8-8,8-8 9-9,9-9 10-10,10-10 11-11,11-11 12-12,12-12
as i have little time , i am immediately in medias res . | 0-0,0-0 1-1,1-1 2-2,4-4 3-4,2-3 5-5,5-5 6-6,7-7 7-7,6-6 8-8,8-8 9-9,9-9 10-10,10-10 11-11,11-11 12-12,12-12
as i have little time , i would immediately in medias res . | 0-1,0-1 2-2,4-4 3-4,2-3 5-5,5-5 6-6,7-7 7-7,6-6 8-8,8-8 9-9,9-9 10-10,10-10 11-11,11-11 12-12,12-12
because i have little time , i am immediately in medias res . | 0-0,0-0 1-1,1-1 2-2,4-4 3-4,2-3 5-5,5-5 6-7,6-7 8-8,8-8 9-9,9-9 10-10,10-10 11-11,11-11 12-12,12-12
as i have little time , i am immediately in medias res . | 0-1,0-1 2-2,4-4 3-3,2-2 4-4,3-3 5-5,5-5 6-6,7-7 7-7,6-6 8-8,8-8 9-9,9-9 10-10,10-10 11-11,11-11 12-12,12-12
as i have little time , i am immediately in medias res . | 0-0,0-0 1-1,1-1 2-2,4-4 3-3,2-2 4-4,3-3 5-5,5-5 6-6,7-7 7-7,6-6 8-8,8-8 9-9,9-9 10-10,10-10 11-11,11-11 12-12,12-12
as i have little time , i am in res medias immediately . | 0-1,0-1 2-2,4-4 3-4,2-3 5-5,5-5 6-7,6-7 8-8,9-9 9-9,11-11 10-10,10-10 11-11,8-8 12-12,12-12
because i have little time , i am immediately in medias res . | 0-1,0-1 2-2,4-4 3-4,2-3 5-5,5-5 6-7,6-7 8-8,8-8 9-9,9-9 10-10,10-10 11-11,11-11 12-12,12-12
as i have little time , i am in res medias immediately . | 0-0,0-0 1-1,1-1 2-2,4-4 3-4,2-3 5-5,5-5 6-7,6-7 8-8,9-9 9-9,11-11 10-10,10-10 11-11,8-8 12-12,12-12