

Phrase-based Statistical Machine Translation

Ulrich Germann

September 10, 2014



Decoding in Phrase-based SMT

“Decoding” is SMT lingo for “translation”

November inflation rates were higher than expected in the 13 countries of the eurozone .

Decoding in Phrase-based SMT

“Decoding” is SMT lingo for “translation”


November inflation rates were higher than expected in the 13 countries of the euro. . .



<s> Teuerungsraten
Inflationsraten
...

Decoding in Phrase-based SMT

“Decoding” is SMT lingo for “translation”


November inflation rates were higher than expected in the 13 countries of the euro. . .

<s> Inflationsraten

Decoding in Phrase-based SMT

“Decoding” is SMT lingo for “translation”

November inflation rates were higher than expected in the 13 countries of the euro. . .

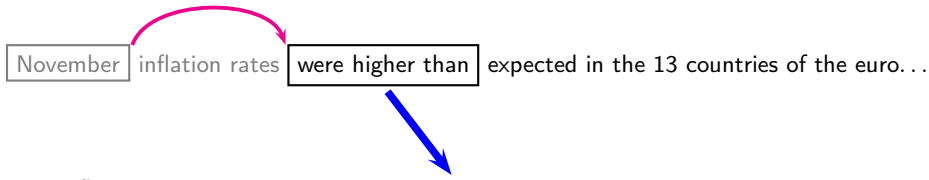


<s> Inflationsraten im November

Decoding in Phrase-based SMT

“Decoding” is SMT lingo for “translation”

November inflation rates were higher than expected in the 13 countries of the euro. . .



⟨s⟩ Inflationsraten im November waren höher als

Decoding in Phrase-based SMT

“Decoding” is SMT lingo for “translation”

November inflation rates were higher than expected in the 13 countries of the euro. . .

⟨s⟩ Inflationsraten im November waren höher als erwartet in den

Decoding in Phrase-based SMT

“Decoding” is SMT lingo for “translation”

November inflation rates were higher than expected in the 13 countries of the euro...

⟨s⟩ Inflationraten im November waren höher als erwartet in den **13 Ländern**

Decoding in Phrase-based SMT

“Decoding” is SMT lingo for “translation”

... inflation rates were higher than expected in the 13 countries of the eurozone .

⟨s⟩ Inflationraten ... waren höher als erwartet in den 13 Ländern der Eurozone .

Scoring Translation Hypotheses in PBSMT

Log-linear combination of:

Translation Model assesses the quality of phrase-level translations.

Distortion Model evaluates jumps between source phrases.

Language Model evaluates the fluency of the translation hypothesis

$$P(\textit{translation} | \textit{source}) = \exp \left(\begin{array}{l} \alpha_{TM} \log P_{TM}(\textit{translation} | \textit{source}) \\ + \alpha_{DM} \log P_{DM}(\textit{translation} | \textit{source}) \\ + \alpha_{LM} \log P_{LM}(\textit{translation} | \textit{source}) \end{array} \right)$$

Scoring Translation Hypotheses in PBSMT

Log-linear combination of:

- Translation Model** assesses the quality of phrase-level translations.
- Distortion Model** evaluates jumps between source phrases.
- Language Model** evaluates the fluency of the translation hypothesis

November inflation rates were higher than expected in the 13 countries of the eurozone .

Scoring Translation Hypotheses in PBSMT

Log-linear combination of:

- Translation Model** assesses the quality of phrase-level translations.
- Distortion Model** evaluates jumps between source phrases.
- Language Model** evaluates the fluency of the translation hypothesis

November inflation rates were higher than expected in the 13 countries of the euro. . .



Teuerungsraten
(s) **Inflationsraten**

...
 $p(t | i, \mathcal{M}_{tr}, \mathcal{M}_{lm}, \mathcal{M}_d) =$

$$\exp \left(\alpha_{tr} \cdot \log p_{tr} (\text{Inflationsraten} | \text{inflation rates}) + \alpha_{lm} \cdot \log p_{lm} (\text{Inflationsraten} | \langle s \rangle) \right)$$

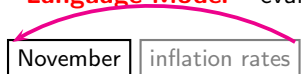
Scoring Translation Hypotheses in PBSMT

Log-linear combination of:

Translation Model assesses the quality of phrase-level translations.

Distortion Model evaluates jumps between source phrases.

Language Model evaluates the fluency of the translation hypothesis

 November | inflation rates | were higher than expected in the 13 countries of the euro. . .

$\langle s \rangle$ Inflationsraten

$p(t | i, \mathcal{M}_{tr}, \mathcal{M}_{lm}, \mathcal{M}_d) =$

$$\exp \left(\begin{array}{l} \alpha_{tr} \cdot \log p_{tr}(\text{Inflationsraten} | \text{inflation rates}) + \alpha_{lm} \cdot \log p_{lm}(\text{Inflationsraten} | \langle s \rangle) \\ + \alpha_d \cdot \log p_d(-2) \end{array} \right)$$

Scoring Translation Hypotheses in PBSMT

Log-linear combination of:

- Translation Model** assesses the quality of phrase-level translations.
- Distortion Model** evaluates jumps between source phrases.
- Language Model** evaluates the fluency of the translation hypothesis

November inflation rates were higher than expected in the 13 countries of the euro. . .


<s> Inflationraten im November

$p(t | i, \mathcal{M}_{tr}, \mathcal{M}_{lm}, \mathcal{M}_d) =$

$$\exp \left(\begin{array}{ll} \alpha_{tr} \cdot \log p_{tr}(\text{Inflationraten} | \text{inflation rates}) & + \alpha_{lm} \cdot \log p_{lm}(\text{Inflationraten} | \langle s \rangle) \\ + \alpha_d \cdot \log p_d(-2) & \\ + \alpha_{tr} \cdot \log p_{tr}(\text{im November} | \text{November}) & + \alpha_{lm} \cdot \log p_{lm}(\text{im} | \dots \text{Inflationraten}) \\ & + \alpha_{lm} \cdot \log p_{lm}(\text{November} | \dots \text{in}) \end{array} \right)$$

Scoring Translation Hypotheses in PBSMT

Log-linear combination of:

- Translation Model** assesses the quality of phrase-level translations.
- Distortion Model** evaluates jumps between source phrases.
- Language Model** evaluates the fluency of the translation hypothesis

November inflation rates were higher than expected in the 13 countries of the euro...

(s) Inflationsraten im November waren höher als

$$p(t | i, \mathcal{M}_{tr}, \mathcal{M}_{lm}, \mathcal{M}_d) =$$

$$\exp \left(\begin{array}{ll} \alpha_{tr} \cdot \log p_{tr}(\text{Inflationsraten} | \text{inflation rates}) & + \alpha_{lm} \cdot \log p_{lm}(\text{Inflationsraten} | \langle s \rangle) \\ + \alpha_d \cdot \log p_d(-2) & \\ + \alpha_{tr} \cdot \log p_{tr}(\text{im November} | \text{November}) & + \alpha_{lm} \cdot \log p_{lm}(\text{im} | \dots \text{Inflationsraten}) \\ & + \alpha_{lm} \cdot \log p_{lm}(\text{November} | \dots \text{in}) \\ + \alpha_d \cdot \log p_d(+3) & \\ + \alpha_{tr} \cdot \log p_{tr}(\text{waren} \dots \text{als} | \text{were} \dots \text{than}) & + \alpha_{lm} \cdot \log p_{lm}(\text{waren} | \dots \text{November}) \\ & + \alpha_{lm} \cdot \log p_{lm}(\text{höher} | \dots \text{waren}) \\ & + \alpha_{lm} \cdot \log p_{lm}(\text{als} | \dots \text{höher}) \end{array} \right)$$

Scoring Translation Hypotheses in PBSMT

Log-linear combination of:

- Translation Model** assesses the quality of phrase-level translations.
- Distortion Model** evaluates jumps between source phrases.
- Language Model** evaluates the fluency of the translation hypothesis

November inflation rates were higher than expected in the 13 countries of the euro...

⟨s⟩ Inflationraten im November waren höher als **erwartet in den**

$$p(t | i, \mathcal{M}_{tr}, \mathcal{M}_{lm}, \mathcal{M}_d) =$$

$$\exp \left(\begin{array}{ll} \alpha_{tr} \cdot \log p_{tr}(\text{Inflationraten} | \text{inflation rates}) & + \alpha_{lm} \cdot \log p_{lm}(\text{Inflationraten} | \langle s \rangle) \\ + \alpha_d \cdot \log p_d(-2) & \\ + \alpha_{tr} \cdot \log p_{tr}(\text{im November} | \text{November}) & + \alpha_{lm} \cdot \log p_{lm}(\text{im} | \dots \text{Inflationraten}) \\ & + \alpha_{lm} \cdot \log p_{lm}(\text{November} | \dots \text{in}) \\ + \alpha_d \cdot \log p_d(+3) & \\ + \alpha_{tr} \cdot \log p_{tr}(\text{waren} \dots \text{als} | \text{were} \dots \text{than}) & + \alpha_{lm} \cdot \log p_{lm}(\text{waren} | \dots \text{November}) \\ & + \alpha_{lm} \cdot \log p_{lm}(\text{höher} | \dots \text{waren}) \\ & + \alpha_{lm} \cdot \log p_{lm}(\text{als} | \dots \text{höher}) \end{array} \right)$$

Scoring Translation Hypotheses in PBSMT

Log-linear combination of:

- Translation Model** assesses the quality of phrase-level translations.
- Distortion Model** evaluates jumps between source phrases.
- Language Model** evaluates the fluency of the translation hypothesis

November inflation rates were higher than expected in the 13 countries of the euro...

⟨s⟩ Inflationraten im November waren höher als erwartet in den **13 Ländern**

$$p(t | i, \mathcal{M}_{tr}, \mathcal{M}_{lm}, \mathcal{M}_d) =$$

$$\exp \left(\begin{array}{ll} \alpha_{tr} \cdot \log p_{tr}(\text{Inflationraten} | \text{inflation rates}) & + \alpha_{lm} \cdot \log p_{lm}(\text{Inflationraten} | \langle s \rangle) \\ + \alpha_d \cdot \log p_d(-2) & \\ + \alpha_{tr} \cdot \log p_{tr}(\text{im November} | \text{November}) & + \alpha_{lm} \cdot \log p_{lm}(\text{im} | \dots \text{Inflationraten}) \\ & + \alpha_{lm} \cdot \log p_{lm}(\text{November} | \dots \text{in}) \\ + \alpha_d \cdot \log p_d(+3) & \\ + \alpha_{tr} \cdot \log p_{tr}(\text{waren} \dots \text{als} | \text{were} \dots \text{than}) & + \alpha_{lm} \cdot \log p_{lm}(\text{waren} | \dots \text{November}) \\ & + \alpha_{lm} \cdot \log p_{lm}(\text{höher} | \dots \text{waren}) \\ & + \alpha_{lm} \cdot \log p_{lm}(\text{als} | \dots \text{höher}) \end{array} \right)$$

Scoring Translation Hypotheses in PBSMT

Log-linear combination of:

- Translation Model** assesses the quality of phrase-level translations.
- Distortion Model** evaluates jumps between source phrases.
- Language Model** evaluates the fluency of the translation hypothesis

... inflation rates were higher than expected in the 13 countries of the eurozone .

⟨s⟩ Inflationraten ... waren höher als erwartet in den 13 Ländern **der Eurozone** .

$$p(t | i, \mathcal{M}_{tr}, \mathcal{M}_{lm}, \mathcal{M}_d) =$$

$$\exp \left(\begin{array}{ll} \alpha_{tr} \cdot \log p_{tr}(\text{Inflationraten} | \text{inflation rates}) & + \alpha_{lm} \cdot \log p_{lm}(\text{Inflationraten} | \langle s \rangle) \\ + \alpha_d \cdot \log p_d(-2) & \\ + \alpha_{tr} \cdot \log p_{tr}(\text{im November} | \text{November}) & + \alpha_{lm} \cdot \log p_{lm}(\text{im} | \dots \text{Inflationraten}) \\ & + \alpha_{lm} \cdot \log p_{lm}(\text{November} | \dots \text{in}) \\ + \alpha_d \cdot \log p_d(+3) & \\ + \alpha_{tr} \cdot \log p_{tr}(\text{waren} \dots \text{als} | \text{were} \dots \text{than}) & + \alpha_{lm} \cdot \log p_{lm}(\text{waren} | \dots \text{November}) \\ & + \alpha_{lm} \cdot \log p_{lm}(\text{höher} | \dots \text{waren}) \\ & + \alpha_{lm} \cdot \log p_{lm}(\text{als} | \dots \text{höher}) \end{array} \right)$$

Distortion Modeling

- Exponential probability decay over distance:

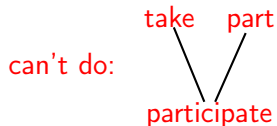
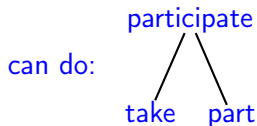
$$p_d(x) = \gamma^{\text{abs}(x)}$$

- Lexicalized discrete model (Koehn et al., 2005)
 - Estimated separately for each phrase.
 - Three types of *type(j)* of jumps:
 - mono** phrase immediately follows the previously translated phrase
 - swap** phrase swaps positions with the previously translated phrase
 - other** anything else
- ...

Building Phrase tables

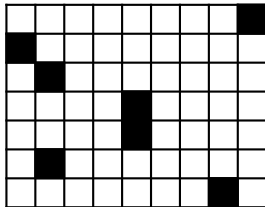
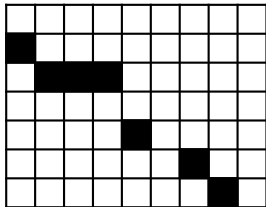
recap from Monday ...

- IBM models for word alignment
- let's skip the math ... \Rightarrow one-to-many alignments:
 - each target word is aligned with at most one source word
 - each source word can be aligned with several target words



Alignment Symmetrisation

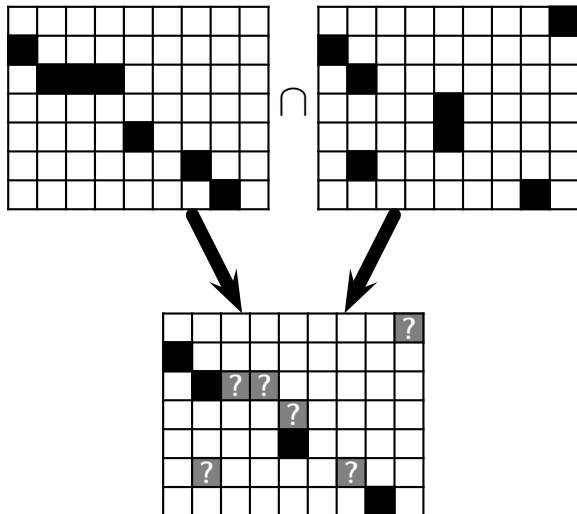
grow-diag + final-and



Alignment Symmetrisation

grow-diag + final-and

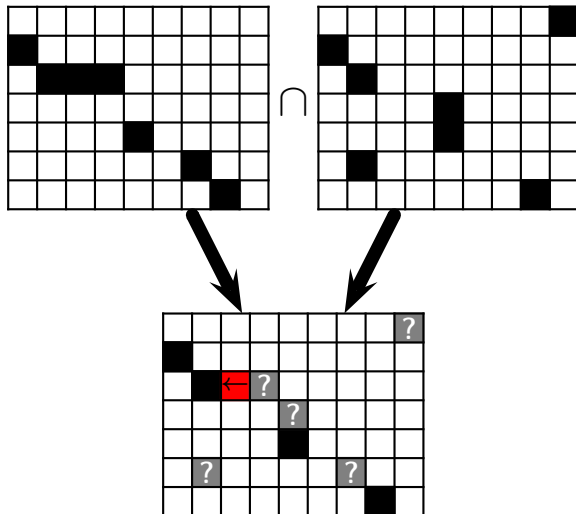
Step 1: Intersect the two alignments:



Alignment Symmetrisation

grow-diag + final-and

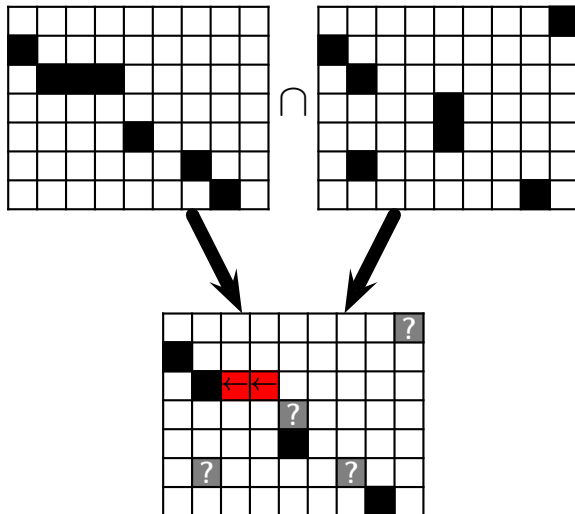
Step 1: Intersect the two alignments:



Alignment Symmetrisation

grow-diag + final-and

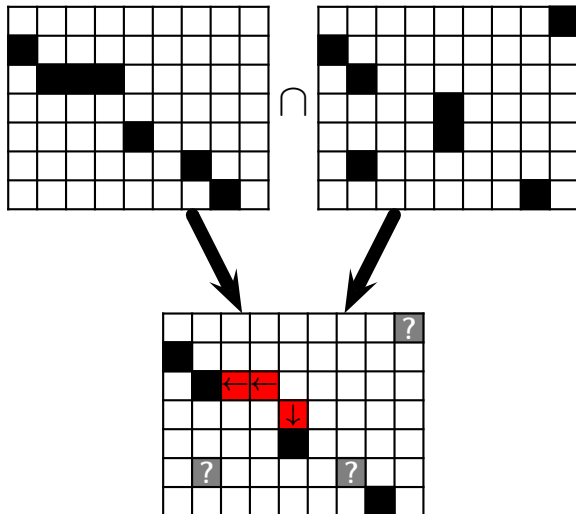
Step 1: Intersect the two alignments:



Alignment Symmetrisation

grow-diag + final-and

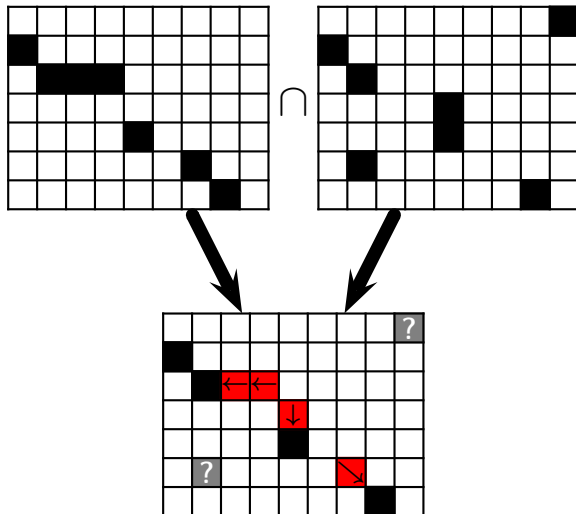
Step 1: Intersect the two alignments:



Alignment Symmetrisation

grow-diag + final-and

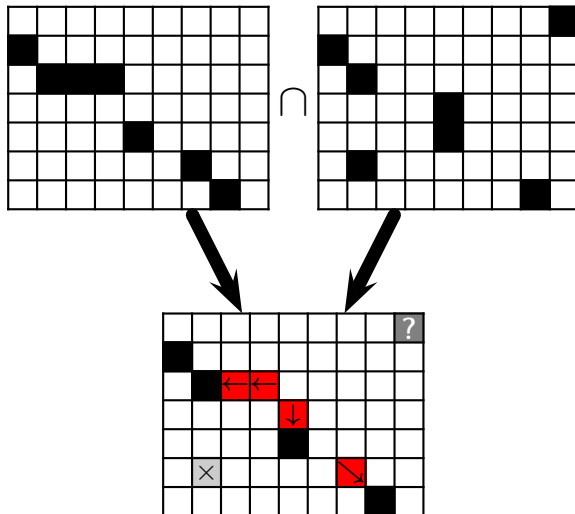
Step 1: Intersect the two alignments:



Alignment Symmetrisation

grow-diag + final-and

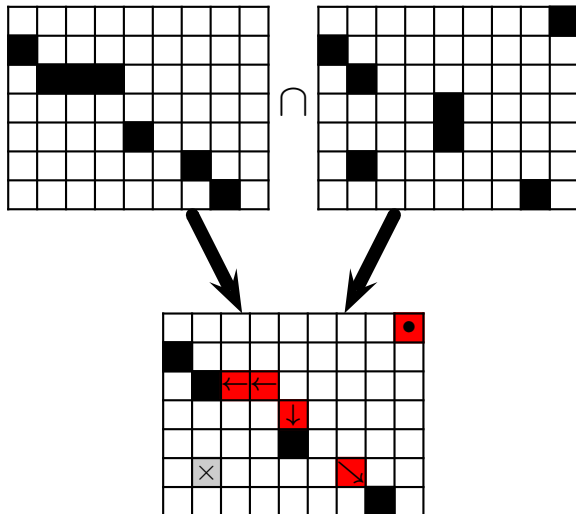
Step 1: Intersect the two alignments:



Alignment Symmetrisation

grow-diag + final-and

Step 1: Intersect the two alignments:



Building Phrase Tables

phrase extraction

	My	question	relates	to	something	that	will	come	up	on	Thursday	and	which	will	then	raise	again	.	
Meine	■																		
Frage		■																	
betrifft			■	■															
eine					■														
Angelegenheit						■													
,							■												
die								■											
am									■										
Donnerstag										■									
zur											■								
Sprache												■	■						
kommen														■	■				
wird																■			
und																	■		
auf																		■	
die																			■
ich																			■
dann																			■
erneut																			■
verweisen																			■
werde																			■
.																			■

Phrase Table

meine ⇔ my
Frage ⇔ question

Building Phrase Tables

phrase extraction

	My	question	relates	to	something	that	will	come	up	on	Thursday	and	which	will	then	raise	again	.
Meine	■	■																
Frage	■	■																
betrifft			■	■														
eine					■	■												
Angelegenheit						■												
,																		
die						■												
am										■								
Donnerstag											■							
zur												■	■					
Sprache												■	■					
kommen												■	■					
wird														■				
und																		
auf																		
die																		
ich																		
dann																		
erneut																		
verweisen																		
werde																		
.																		

Phrase Table

meine ⇔ my
meine Frage ⇔ my question

Building Phrase Tables

phrase extraction

	My	question	relates	to	something	that	will	come	on	Thursday	and	which	will	then	raise	again	.
Meine	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
Frage	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
betrifft	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
eine	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
Angelegenheit	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
,	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
die	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
am	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
Donnerstag	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
zur	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
Sprache	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
kommen	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
wird	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
und	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
auf	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
die	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
ich	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
dann	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
erneut	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
verweisen	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
werde	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■
.	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■

Phrase Table

meine ⇔ my
meine Frage ⇔ my question
meine Frage betrifft ⇔ my question relates to

Building Phrase Tables

phrase extraction

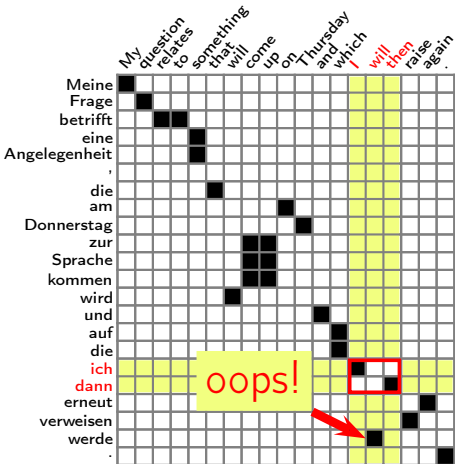
	My	question	relates	to	something	that	will	come	up	on	Thursday	and	which	will	then	raise	again	.	
Meine	■																		
Frage		■																	
betrifft			■																
eine				■															
Angelegenheit					■														
,																			
die					■														
am										■									
Donnerstag											■								
zur												■							
Sprache												■							
kommen												■							
wird													■						
und														■					
auf															■				
die																■			
ich																	■		
dann																		■	
erneut																			■
verweisen																			■
werde																			■
.																			■

Phrase Table

meine	↔	my
meine Frage	↔	my question
meine Frage betrifft	↔	my question relates to
meine Frage betrifft eine Angelegenheit	↔	my question relates to something

Building Phrase Tables

phrase extraction



Phrase Table

meine	↔	my
meine Frage	↔	my question
meine Frage betrifft	↔	my question relates to
meine Frage betrifft	↔	my question relates to
eine Angelegenheit	↔	something
.		.
.		.
Frage	↔	question
Frage betrifft	↔	question relates to
.		.
.		.
ich dann	↔	I will then
.		.
.		.

Scoring Phrase Table Entries

[again:] weighted linear combination of features:

$$P_{TM}(t | s) = \exp \left(\sum_j \alpha_j f_j(s, t) \right)$$

Scoring Phrase Table Entries: Feature Functions

- log of smoothed **forward** cond. prob.:

$$\text{smooth} \left(\frac{\text{count}(\text{target phrase})}{\text{count}(\text{source phrase})} \right)$$

- log of smoothed **backward** cond. prob.:

$$\text{smooth} \left(\frac{\text{count}(\text{source phrase})}{\text{count}(\text{target phrase})} \right)$$

- “lexically smoothed” (Zens&Ney) **forward** probability

$$\sum_t \log P(t \mid \text{source phrase}, \text{alignment})$$

- “lexically smoothed” **backward** probability

$$\sum_s \log P(s \mid \text{target phrase}, \text{alignment})$$

- length of target phrase (“word penalty”)
- 1 (“phrase penalty”)

Decoding

based on slides originally by P. Koehn, edited by M. Huck (and possibly others)

Given the model, find the best translation

$$\mathbf{e}_{\text{best}} = \operatorname{argmax}_{\mathbf{e}} p(\mathbf{e} | \mathbf{f})$$

We use the “Viterbi approximation”

$$(a, \mathbf{e})_{\text{best}} = \operatorname{argmax}_{(a, \mathbf{e})} p(a, \mathbf{e} | \mathbf{f})$$

- This is a search problem - a big one.
 - Dynamic programming
 - Approximation (beam search)
 - Model restrictions (reordering)

Decoding

Maria	no	dio	una	bofetada	a	la	bruja	verde
-------	----	-----	-----	----------	---	----	-------	-------

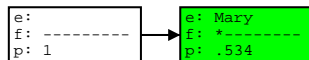
Mary not give a slap to the witch green
did not a slap by green witch
no slap to the
did not give to
the
slap the witch

- many different ways to *segment* the input sentence into phrases
- many different ways to *translate* each phrase

Hypothesis Expansion

María	no	dio	una	bofetada	a	la	bruja	verde
-------	----	-----	-----	----------	---	----	-------	-------

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		

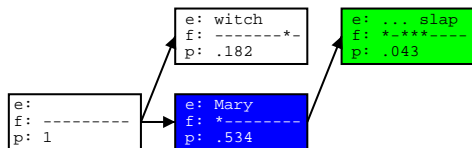


- Pick *translation option*
- Create *hypothesis*
 - e: add English phrase Mary
 - f: first foreign word covered
 - p: probability 0.534

Hypothesis Expansion

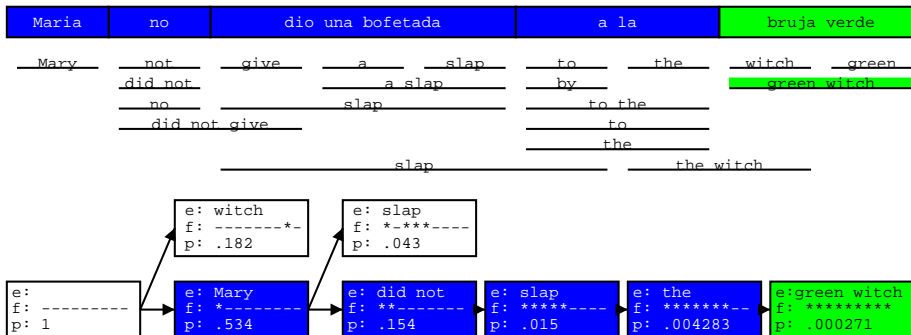
Maria	no	dio una bofetada	a	la	bruja	verde
-------	----	------------------	---	----	-------	-------

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



- Further hypothesis expansion

Hypothesis Expansion

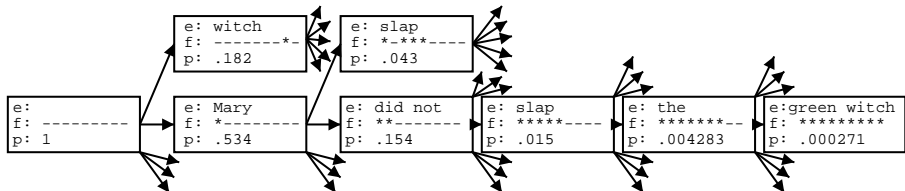


- ...until all foreign words covered
 - find *best hypothesis* that covers all foreign words
 - *backtrack* to read off translation

Hypothesis Expansion

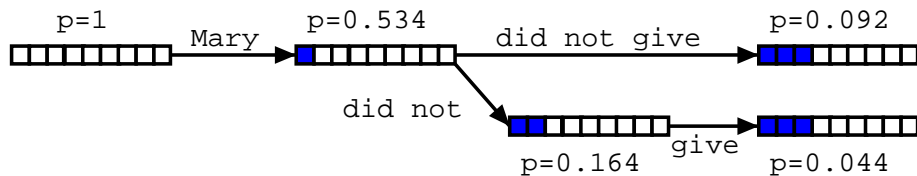
Maria	no	dio	una	bofetada	a	la	bruja	verde
-------	----	-----	-----	----------	---	----	-------	-------

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



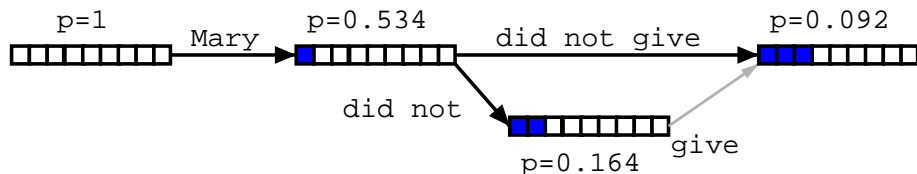
Adding more hypothesis \Rightarrow Explosion of search space

Hypothesis Recombination



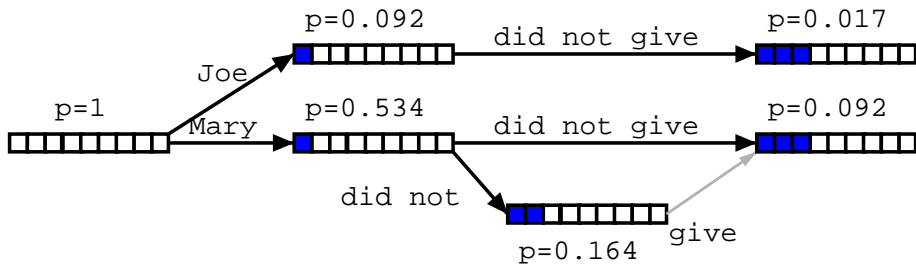
- Different paths to the *same* partial translation

Hypothesis Recombination



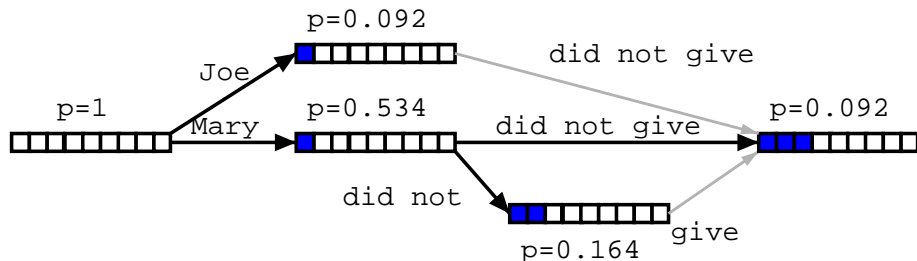
- Different paths to the same partial translation
- ⇒ *Combine paths*
- *drop weaker path*
 - keep pointer from weaker path (for lattice generation)

Hypothesis Recombination



- Recombined hypotheses do *not* have to *match completely*
- No matter what is added, weaker path can be dropped, if:
 - *last $n - 1$ English words match* (matters for language model)
 - *foreign word coverage vectors match* (affects future path)

Hypothesis Recombination



- Recombined hypotheses do not have to match completely
- No matter what is added, weaker path can be dropped, if:
 - last $n - 1$ English words match (matters for language model)
 - foreign word coverage vectors match (effects future path)

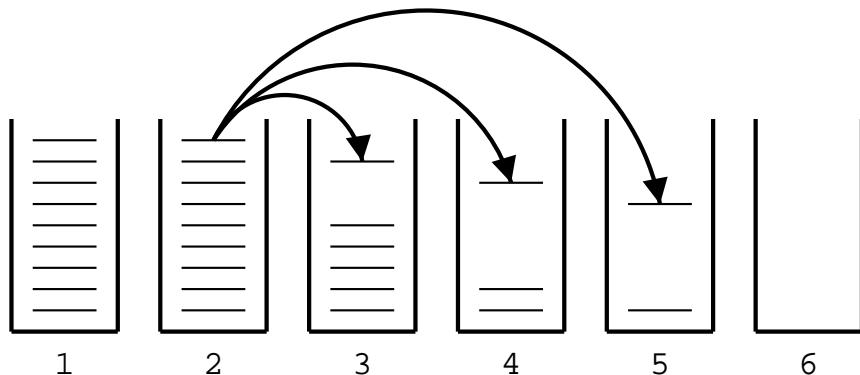
⇒ *Combine paths*

Beam Search

heuristically *discard* weak hypotheses early

- it is better to organize hypotheses in stacks (actually: priority queues), e.g. by
 - *same* foreign words covered
 - *same number* of foreign words covered
- compare hypotheses in stacks, discard bad ones
 - **histogram pruning**: keep top k hypotheses in each stack (e.g., $k=100$)
 - **threshold pruning**: keep hypotheses that are at most α times the cost of best hypothesis in stack (e.g., $\alpha = 0.001$)

Hypothesis Stacks

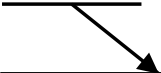


- Organization of hypotheses into stacks
 - here: based on *number of foreign words* translated
 - during translation all hypotheses from one stack are expanded
 - expanded hypotheses are placed into stacks

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**



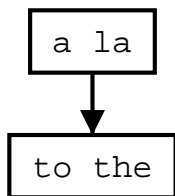
```
e: the
f: -----**--
p: 0.354
```

**covers
easier part
--> lower cost**

- Hypothesis that covers *easy part* of sentence is preferred
- ⇒ Need to consider **future cost** of uncovered parts

Future Cost Estimation

Step 1: estimate future cost for each *translation option*

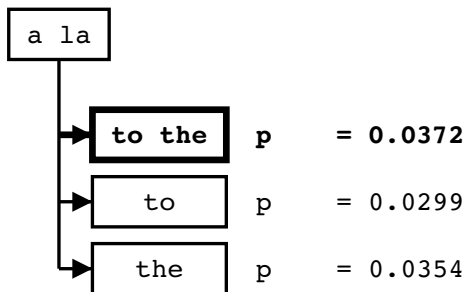


- look up translation model cost
- estimate language model cost (no prior context)
- ignore reordering model cost

$$\Rightarrow \text{LM} * \text{TM} = p(\text{to}) * p(\text{the}|\text{to}) * p(\text{to the}|\text{a la})$$

Future Cost Estimation

Step 2: find *cheapest cost* (highest probability) among translation options

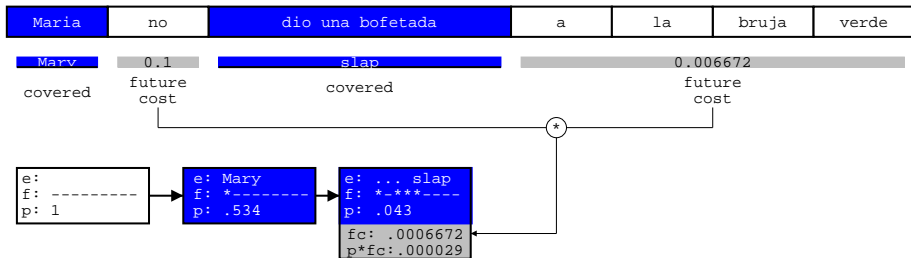


Future Cost Estimation

Step 3: Find *lowest future cost* for each possible span

- Cost of translation option for that span, *or*
 - Sum of costs of covering subspans
- ⇒ Pre-compute future costs, bottom up., via dynamic programming.

Future Cost Estimation: Application

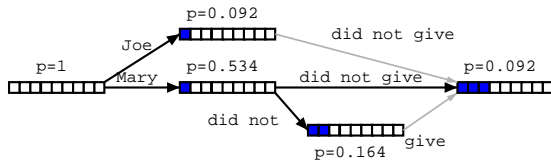


- Use future cost estimates when *pruning* hypotheses
- For each *uncovered continuous span*:
 - look up *future costs* for each maximal contiguous uncovered span
 - *add* to actually accumulated cost for translation option for pruning

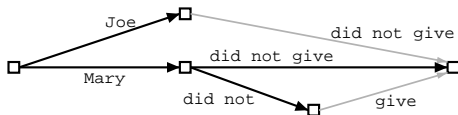
Limits on Reordering

- Reordering may be **limited**
 - **Monotone** translation: No reordering at all
 - Only phrase movements of at most d words
- Reordering limits *speed* up search (polynomial instead of exponential)
- Current reordering models are weak, so limits *improve* translation quality

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



Sample N-Best List

- Simple N-best list:

Translation	Reordering	LM	TM	WordPenalty	Score
this is a small house	0	-27.0908	-1.83258	-5	-28.9234
this is a little house	0	-28.1791	-1.83258	-5	-30.0117
it is a small house	0	-27.108	-3.21888	-5	-30.3268
it is a little house	0	-28.1963	-3.21888	-5	-31.4152
this is an small house	0	-31.7294	-1.83258	-5	-33.562
it is an small house	0	-32.3094	-3.21888	-5	-35.5283
this is an little house	0	-33.7639	-1.83258	-5	-35.5965
this is a house small	-3	-31.4851	-1.83258	-5	-36.3176
this is a house little	-3	-31.5689	-1.83258	-5	-36.4015
it is an little house	0	-34.3439	-3.21888	-5	-37.5628
it is a house small	-3	-31.5022	-3.21888	-5	-37.7211
this is an house small	-3	-32.8999	-1.83258	-5	-37.7325
it is a house little	-3	-31.586	-3.21888	-5	-37.8049
this is an house little	-3	-32.9837	-1.83258	-5	-37.8163
the house is a little	-7	-28.5107	-2.52573	-5	-38.0364
the is a small house	0	-35.6899	-2.52573	-5	-38.2156
is it a little house	-4	-30.3603	-3.91202	-5	-38.2723
the house is a small	-7	-28.7683	-2.52573	-5	-38.294
it 's a small house	0	-34.8557	-3.91202	-5	-38.7677
this house is a little	-7	-28.0443	-3.91202	-5	-38.9563
it 's a little house	0	-35.1446	-3.91202	-5	-39.0566
this house is a small	-7	-28.3018	-3.91202	-5	-39.2139

Summary

- Left-to-right decoding as search
- Hypothesis recombination
- Pruning
- Future cost estimation
- Word lattices and n -best lists