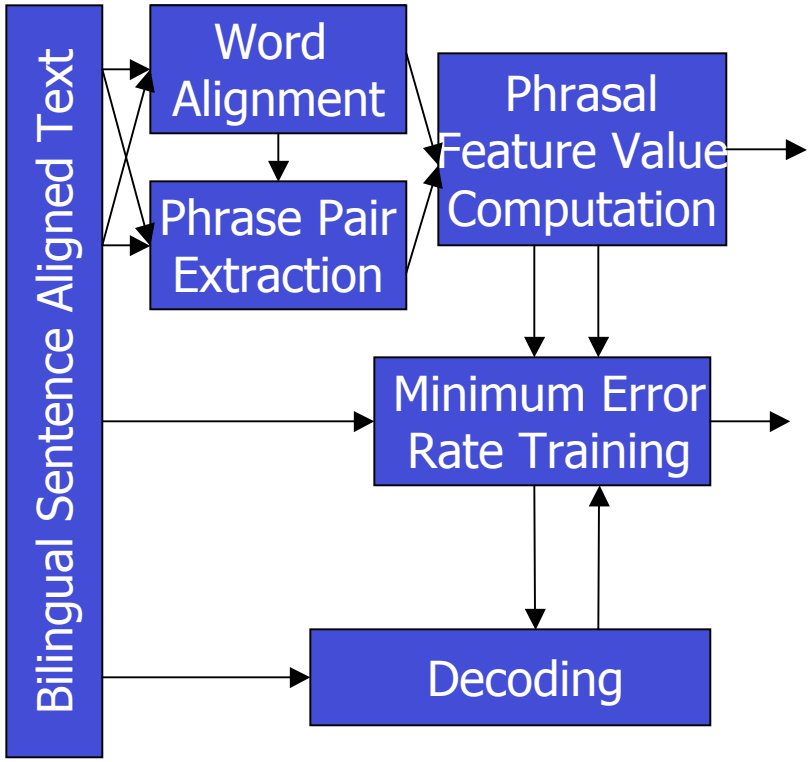
Selective Phrase Pair Extraction for Improved Statistical Machine Translation

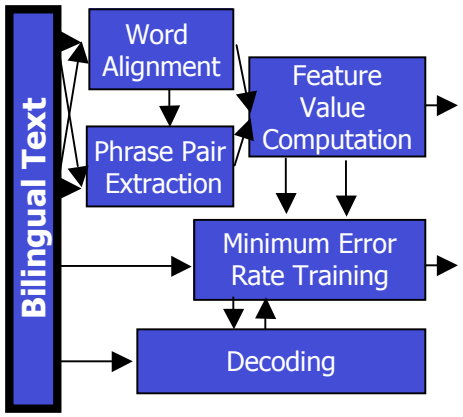
Luke S. Zettlemoyer
MIT CSAIL
and
Robert C. Moore
Microsoft Research

Phrase-based SMT training pipeline



- Many pieces
- We focus on phrase pair extraction component
- First, let's have a quick review of the rest

Bilingual sentence aligned text



je ne parle pas Français
i don't speak French

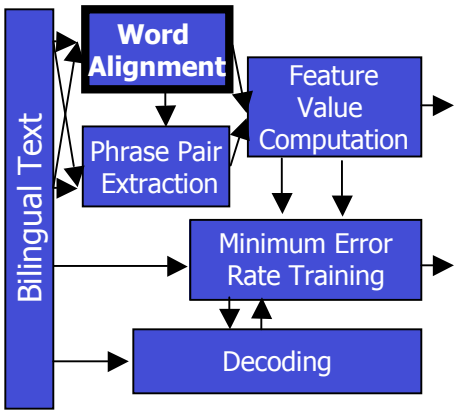
nous acceptons votre opinion
we accept your view

monsieur le Orateur , je invoque le Règlement
Mr. Speaker , I rise on a point of order

…
…

We use Canadian Hansards data in this work.

Word alignment



je ne parle pas Français

i don't speak French

nous acceptons votre opinion

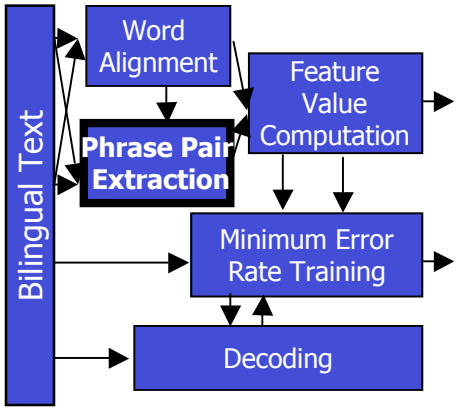
we accept your view

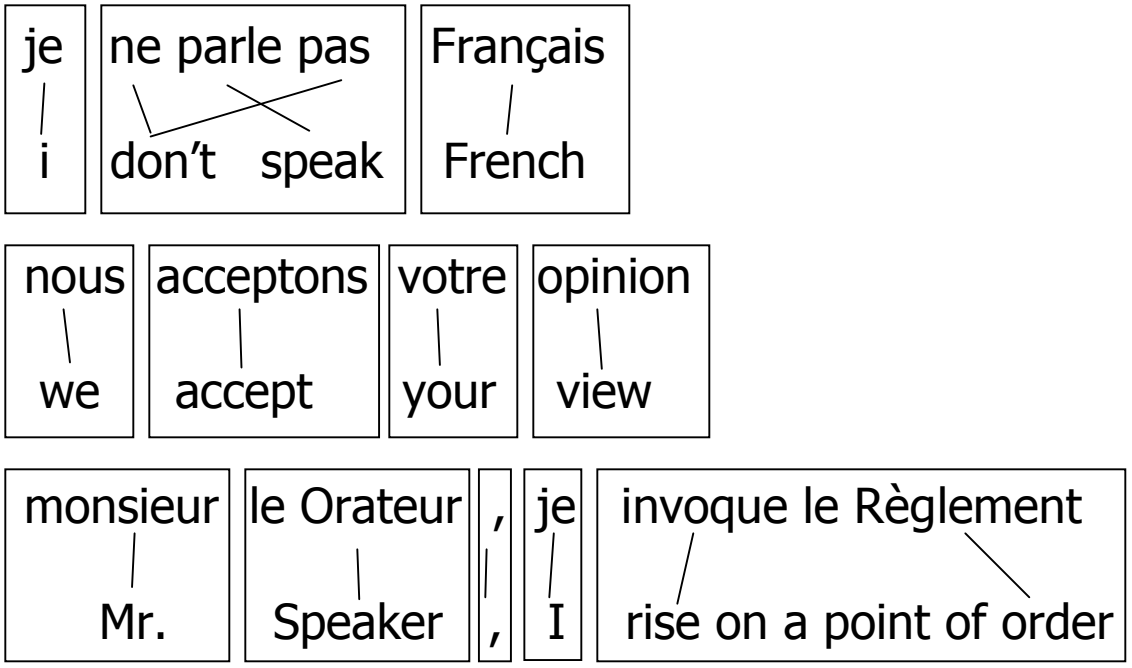
monsieur le Orateur , je invoque le Règlement

Mr. Speaker , I rise on a point of order

See papers by Moore et al. [2005,2006] for more details.

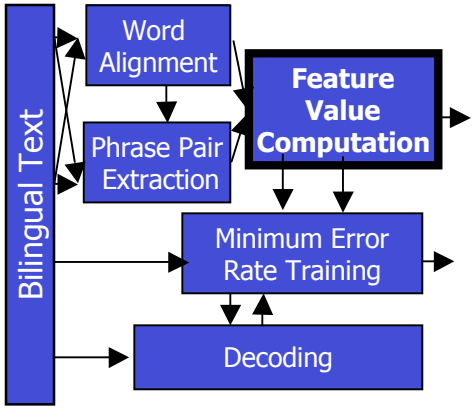
Phrase pair extraction





This step is the focus of the current project.

Phrasal feature value computation



Source Lang. phrase	Target Lang. phrase	log p(s|t)	log p(t|s)	log w(s,t)
je	i	-1.175	-0.776	-0.186
le Orateur	Speaker	-5.522	-0.801	-4.962
nous	we	-0.929	-0.5638	-0.263
monsieur	Mr.	-1.266	-0.01	-1.37
…	…	…	…	…

See paper by Koehn et al. [2003] for more details.

Definitions for phrasal features we use

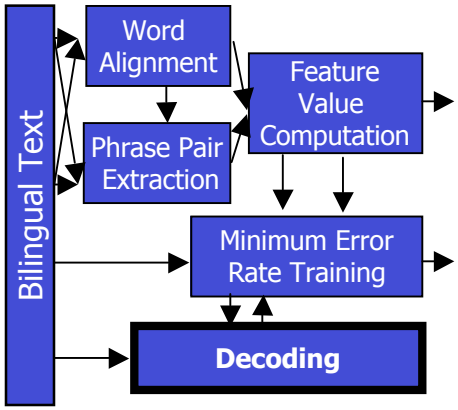
- Translation:
  - *count(s,t)* is the number phrase pairs with source s and target t

$$p(s \mid t) = \frac{count(s,t)}{\sum_{s'} count(s',t)}$$

- Lexical Weighting:
  - n is the length of s
  - m is the length of t
  - $p(s|t)$ is estimated from word aligned corpus

$$w(s,t) = \frac{1}{m} \prod_{i=1}^{n} \sum_{j=1}^{m} p(s_i \mid t_j)$$

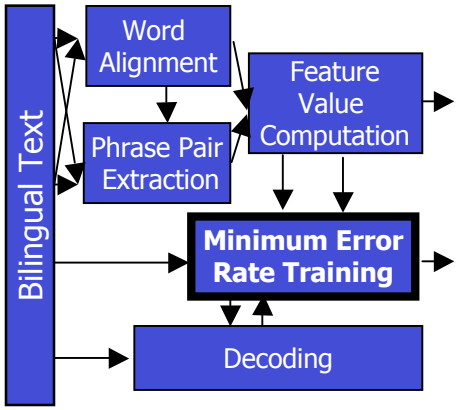
Decoding (translation)



- Searches for highest scoring target sentence for each source sentence
- Uses computed feature values for phrases plus additional features
  - Total number of target sentence words
  - Total number of phrase pairs
  - Distortion penalty
  - N-gram target language model
- We use Koehn's Pharaoh decoder

See Pharaoh manual by Koehn [2004] for more details.

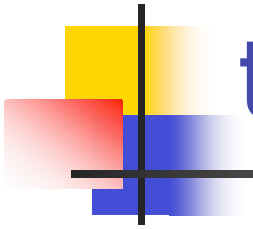
Minimum error rate training



- Repeatedly performs translations to create n-best lists
- Optimize parameters to maximize translation quality (BLEU)
- Output a parameter vector that the decoder will use to translate the test set

See papers by Och et al. [2003, 2004] for more details.

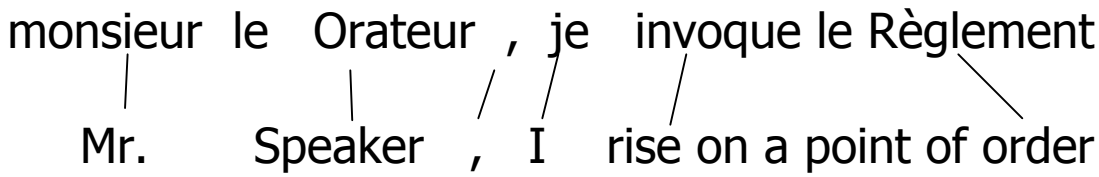
Goal: improve phrase pair table through more selective extraction



- Reduce memory requirements
  - Fewer phrase pairs to store
- Increase translation quality
  - Fewer bad phrase pairs
  - Improved feature values computed for remaining phrase pairs

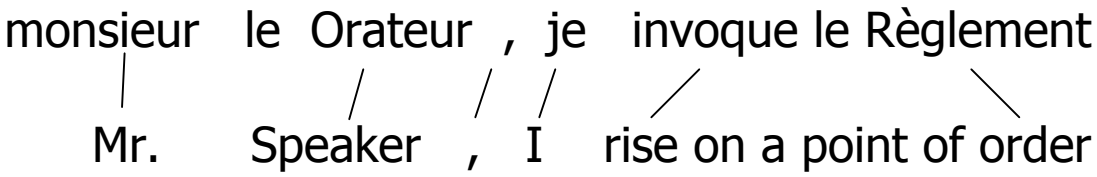
Standard SMT phrase extraction

- Select every possible phrase pair (up to a maximum length) that has at least one word alignment and no crossing word alignments



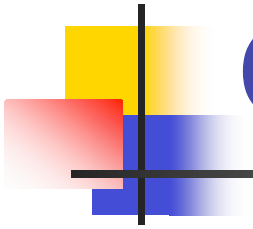
Includes:	
monsieur	Mr.
monsieur le	Mr.
monsieur le Orateur	Mr. Speaker
le Orateur	Speaker
…	…

Does Not Include:	
monsieur le Orateur	Speaker
le Orateur	Mr.
monsieur le	Speaker
le Orateur	Speaker
…	…



All phrases, max length 3:	
monsieur	Mr.
monseiur le	Mr.
monseiur le Orateur	Mr. Speaker
le Oreateur	Speaker
le Oreateur ,	Speaker ,
Orateur	Speaker
Orateur ,	Speaker ,
Orateur , je	Speaker , I
,	,
, je	, I
, je invoque	, I rise
je	I
je invoque	I rise

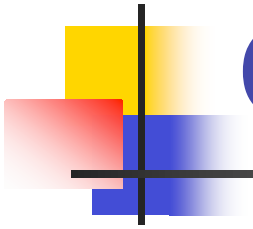
je invoque	I rise on
je invoque le	I rise
je invoque le	I rise on
invoque	rise
invoque	rise on
invoque	rise on a
invoque le	rise
invoque le	rise on
invoque le	rise on a
le Règlement	point of order
le Règlement	of order
le Règlement	order
Règlement	point of order
Règlement	of order
Règlement	order



Our approach

- Standard phrase extraction produces many target language phrases for each source language phrase, and vice versa, due to unaligned words
- Our intuition is that each occurrence of a source or target language phrase really has at most one translation in that occurrence
- So, we try to strictly limit the number of translations selected per phrase occurrence

Our general procedure

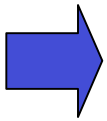


- Perform standard phrase pair extraction
- Compute phrasal feature values and train translation model weights
- Re-extract phrase pairs
  - Select a subset of the original phrase pairs
  - Use sum of phrasal feature values, weighted by translation model weights, to decide which pairs to keep
- Recompute phrasal feature values and retrain translation model weights, using new pair counts

Selecting the phrase pairs

monsieur le Orateur , je invoque le Règlement
Mr. Speaker , I rise on a point of order

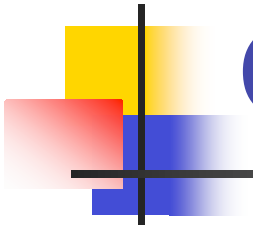
Original phrase pairs with scores:		
monsieur	Mr.	-1
monseiur le	Mr.	-2
le Oreateur	Speaker	-3
Orateur	Speaker	-4
...	...	...
le Règlement	point of order	-100
le Règlement	of order	-101
Règlement	point of order	-102
Règlement	of order	-103



Select some subset of these phrase pairs

Two methods

- Global competitive linking
- Local competitive linking



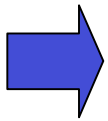
Global competitive linking

- Imposes the global constraint that each phrase is used only once
- For each sentence pair
  - Sort all phrase pairs by their score
  - Select phrase pairs in order of their score, but only if they do not share a phrase with a previously selected pair

Global competitive linking

monsieur le Orateur , je invoque le Règlement
Mr. Speaker , I rise on a point of order

Original phrase pairs with scores:		
monsieur	Mr.	-1
monseiur le	Mr.	-2
le Oreateur	Speaker	-3
Orateur	Speaker	-4
…	…	…
le Règlement	point of order	-100
le Règlement	of order	-101
Règlement	point of order	-102
Règlement	of order	-103

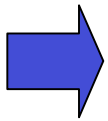


?

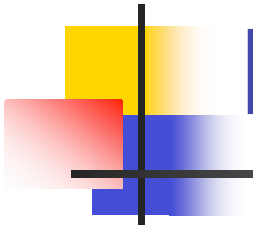
Global competitive linking

monsieur le Orateur , je invoque le Règlement
Mr. Speaker , I rise on a point of order

Original phrase pairs with scores:		
monsieur	Mr.	-1
monseiur le	Mr.	-2
le Oreateur	Speaker	-3
Orateur	Speaker	-4
…	…	…
le Règlement	point of order	-100
le Règlement	of order	-101
Règlement	point of order	-102
Règlement	of order	-103



Selected phrase pairs with scores:		
monsieur	Mr.	-1
~~monseiur le~~	~~Mr.~~	~~-2~~
le Oreateur	Speaker	-3
~~Orateur~~	~~Speaker~~	~~-4~~
…	…	…
le Règlement	point of order	-100
~~le Règlement~~	~~of order~~	~~-101~~
~~Règlement~~	~~point of order~~	~~-102~~
Règlement	of order	-103



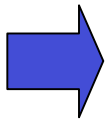
Local competitive linking

- Select the best phrase pair for each source and target language phrase, ignoring global constraints
- For each sentence pair
  - Collect all phrase pairs for a given source or target language phrase
  - Mark the highest scoring pair for each source or target language phrase
  - Select all of the marked phrase pairs

Local competitive linking

monsieur le Orateur , je invoque le Règlement
Mr. Speaker , I rise on a point of order

Original phrase pairs with scores:		
monsieur	Mr.	-1
monseiur le	Mr.	-2
le Oreateur	Speaker	-3
Orateur	Speaker	-4
…	…	…
le Règlement	point of order	-100
le Règlement	of order	-101
Règlement	point of order	-102
Règlement	of order	-103

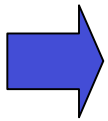


?

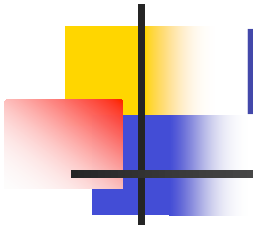
Local competitive linking

monsieur le Orateur , je invoque le Règlement
Mr. Speaker , I rise on a point of order

Original phrase pairs with scores:		
monsieur	Mr.	-1
monseiur le	Mr.	-2
le Oreateur	Speaker	-3
Orateur	Speaker	-4
…	…	…
le Règlement	point of order	-100
le Règlement	of order	-101
Règlement	point of order	-102
Règlement	of order	-103



Selected phrase pairs with scores:		
monsieur	Mr.	-1
monseiur le	Mr.	-2
le Oreateur	Speaker	-3
Orateur	Speaker	-4
…	…	…
le Règlement	point of order	-100
le Règlement	of order	-101
Règlement	point of order	-102
~~Règlement~~	~~of order~~	~~-103~~



Experimental data

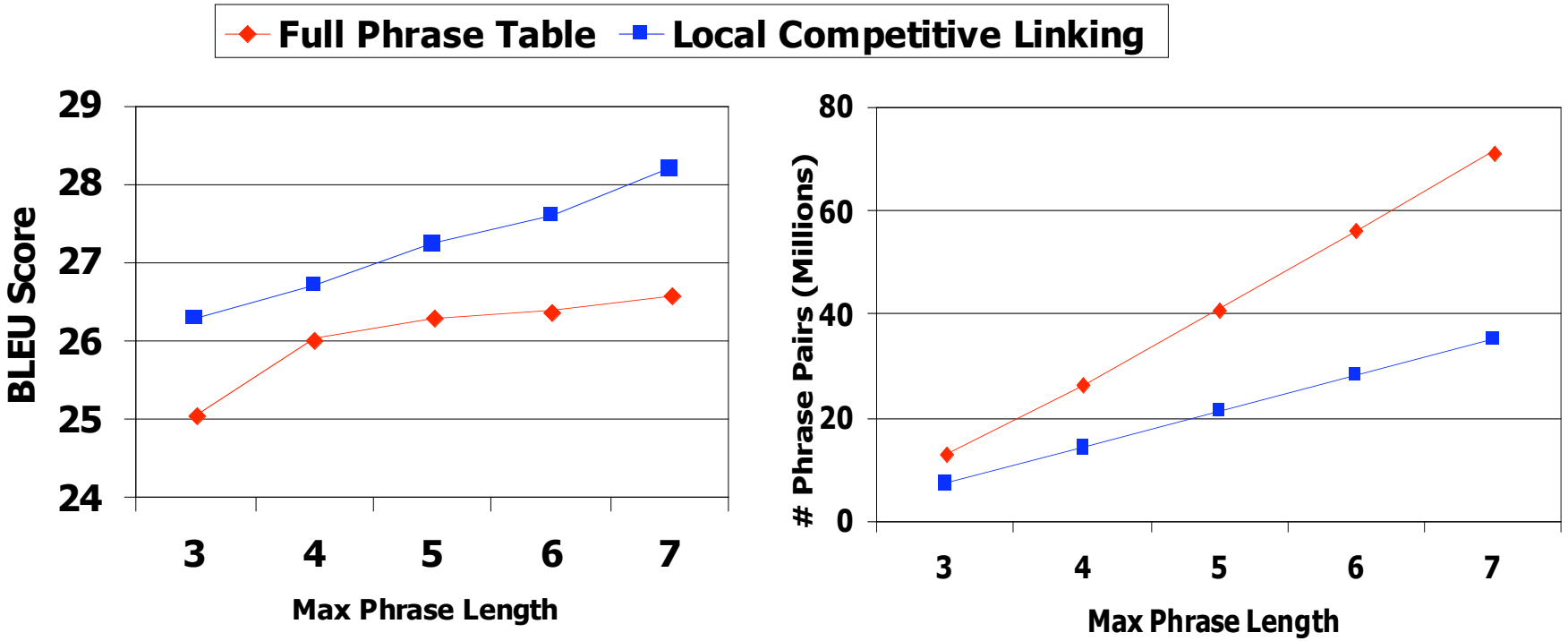
- 500,000 EF Canadian Hansard sentence pairs from 2003 word alignment workshop, word aligned and used for extracting phrase pairs
- Three additional disjoint sets of 2000 sentence pairs from same source used for
  - Training (set translation model weights)
  - Validation (compare selection methods and phrase length limits)
  - Final test

Global vs. local competitive linking

- Phrases up to length 3
- Validation set results:

	# phrase pairs	BLEU
Full phrase pair table	13 M	25.05
Global competitive linking	4.25 M	23.76
Local competitive linking	7.25 M	26.30

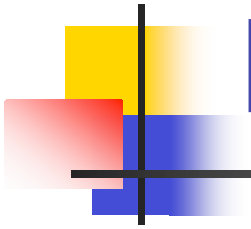
Effect of phrase length limits



Final evaluation

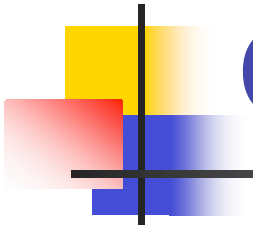
Single run on final test set, using best performing models on validation set (phrases up to length 7)

	BLEU
Full phrase pair table	26.78
Local competitive linking	28.30



Related work

- Other methods have been explored that result in smaller phrase tables than the standard approach [Birch, et al. 2006; De Nero, et al. 2006], but ours seems to be the first that improves BLEU score.
- Phrase table smoothing [Foster et. al., 2006] seems to achieve comparable improvements in BLEU score, but lacks the benefit of reduced memory requirements.



Conclusions and future work

- Selecting phrase pairs according to estimated translation quality
  - Reduces phrase table size
  - Improves BLEU score
- Further research can probably find better ways to
  - Score phrase pairs
  - Use scores to select phrase pairs