

Learning Linear Ordering Problems for Better Translation

Roy Tromble, Google Pittsburgh
Jason Eisner, Johns Hopkins

August 7, 2009

Overview

- Monotonic translation is easier
- Construct sentence-specific Linear Ordering Problems to put German sentences in English word order
- Use reordered sentences to train and decode with another MT system (makes its reordering job easier*)
- Get better translations*

* Hopefully

Deshalb unterstützen und begrüßen **wir** die **vom Haushaltsausschuß** eingebrachten Vorschläge für eine Aufstockung des Haushalts für Prince.

Thus support and welcome **we** the **from the budget committee** tabled proposals for an increase of the budget for Prince.

Deshalb **wir** unterstützen und begrüßen die eingebrachten Vorschläge **vom Haushaltsausschuß** für eine Aufstockung des Haushalts für Prince.

Lassen Sie mich zu einem letzten Punkt **kommen**, der die verbesserte Koordinierung zwischen der Politik der Entwicklungszusammenarbeit und den externen Aspekten der gemeinsamen Fischereipolitik **betrifft**.

Let you me to a last point **come**, which the improved coordination between the policy of the development cooperation and the external aspects of the common fisheries policy **concerns**.

Lassen Sie mich **kommen** zu einem letzten Punkt, der die verbesserte Koordinierung zwischen der Politik **betrifft** der Entwicklungszusammenarbeit und den externen Aspekten der gemeinsamen Fischereipolitik.

“Better Translation”

System	BLEU	TER
Moses	25.55	59.76
This paper	26.40	58.63

- 2000 German-English Europarl sentences
- Significant according to a paired permutation test

Translation is hard

- *Das muß ich erst einmal klären.*
- *That must I for starters resolve.*
- I will have to look into that.

Monotonic translation is easier

- *Das muß ich erst einmal klären.*

- *ich muß erst einmal klären Das.*

- I will have to look into that.

- Prior work: Collins, Koehn, Kučerová (2005)

Reordering automatically

<s>	PDS	VMFIN	PPER	ADV	ADV	VVINF	</s>
Das	muß	ich	erst	einmal	klären		

5

For every pair of words

$$\hat{\pi} = \arg \max_{\pi \in \Pi_n} \sum_{1 \leq i < j \leq n} \theta \cdot \phi(w, \pi_i, \pi_j)$$

The Linear Ordering Problem

	Das	muß	ich	erst	einmal	klären
Das		0.8	0.1	2.0	-0.6	-1.7
muß			-5.4	8.7	8.1	14.8
ich				3.0	3.7	16.0
erst					7.7	1.9
einmal						2.0
klären						

ich muß erst einmal klären Das

Why this model?

- Looks only at the entire source sentence, and reorders for the target language
- Can assign a different score to every permutation
- Can incorporate linguistic information via features
- Can integrate with many translation models

The First difficulty

- The Linear Ordering Problem is NP-complete!
- Proposed solution:
 - Iterate greedy local search
 - with a very large neighborhood
 - using dynamic programming

Local Search using Dynamic Programming

$$S \rightarrow S_{0,n}$$

$$S_{i,k} \rightarrow S_{i,j} S_{j,k}, \max(0, \Delta_{i,j,k})$$

$$S_{i-1,i} \rightarrow \pi_i$$



Iterate!

Das muß ich erst einmal klären

ich muß erst einmal klären Das

Observations

- There are $n!$ permutations to try
- Even the subset we search at each iteration is exponentially large
 - $O(\approx 5.83^n)$
- But search by dynamic programming is fast!
 - $O(n^3)$ with no grammar constant!
- What about the grammar weights?

$$\Delta_{i,j,k} = \sum_{\ell=i+1}^j \sum_{r=j+1}^k B[\pi_r, \pi_\ell] - B[\pi_\ell, \pi_r]$$

Grammar Weights

- Dynamic programming to the rescue!

$$\Delta_{i,j,k} = \sum_{\ell=i+1}^j \sum_{r=j+1}^k B[\pi_r, \pi_\ell] - B[\pi_\ell, \pi_r]$$

	ich	erst	einmal	klären
Das	0.1	2.0	-0.6	-1.7
muß	-5.4	8.7	8.1	14.8

$$\Delta_{i,j,k} = \Delta_{i,j,k-1} + \Delta_{i+1,j,k} - \Delta_{i+1,j,k-1} + B[\pi_k, \pi_{i+1}] - B[\pi_{i+1}, \pi_k]$$

The Second difficulty

- Where do we get the feature weights?
- No previous learning algorithm for the LOP
 - Can't efficiently find the best permutation according to the current model
 - Can't efficiently compute the partition function
- Proposed solution:
 - Approximate perceptron

Learning for the LOP

- Align training sentences
- Compute German' from alignments
- Initialize parameters using log-odds
- Iterative training using greedy local search as an approximation to the true model optimum for the perceptron
- Reordered word pairs:
 - 64% precision
 - 21% recall

Integration with Translation

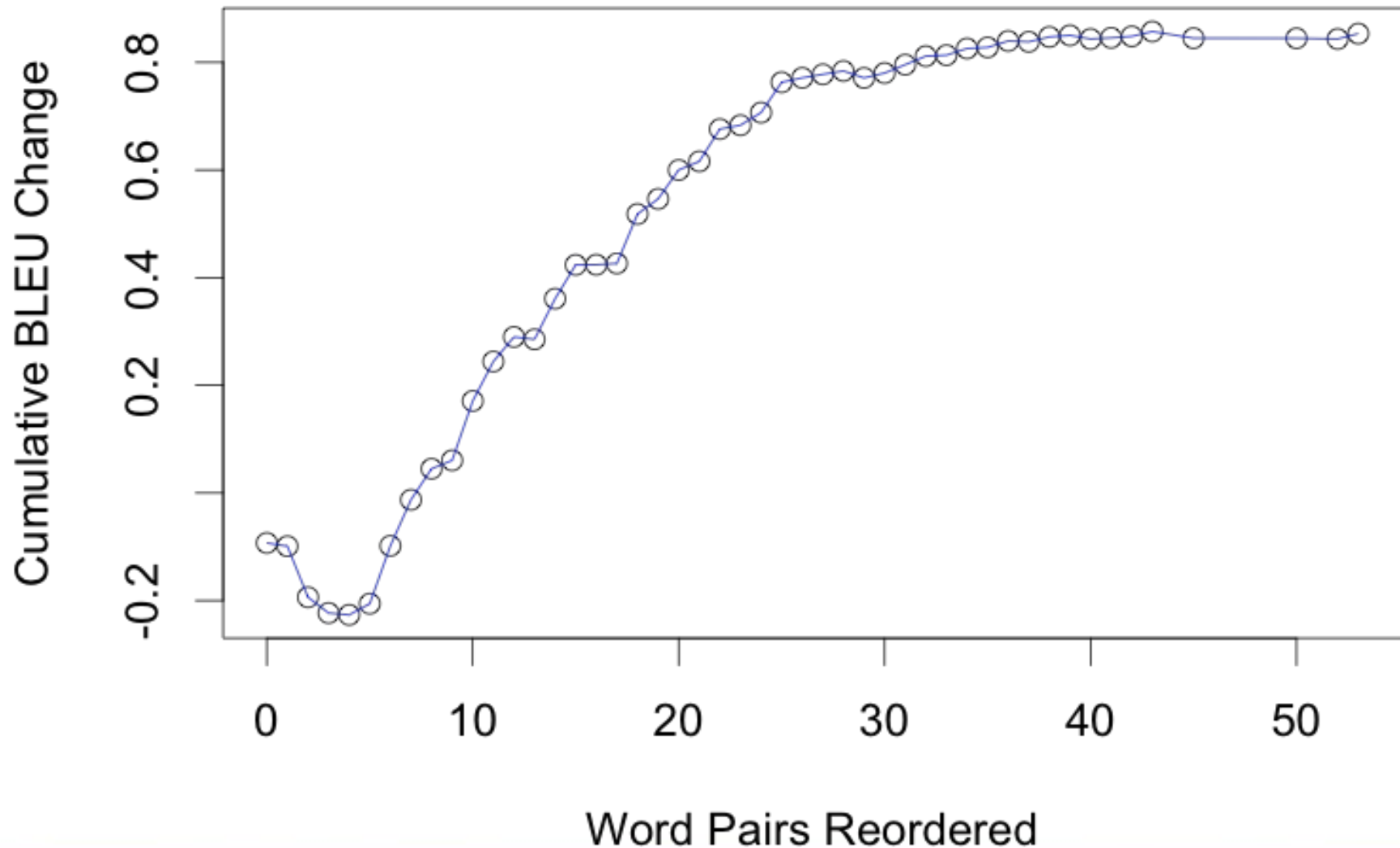
- Reorder the training, dev, and test sentences
- Standard Moses pipeline:
 - Align and learn phrase tables from reordered training sentences
 - Minimum error rate training over reordered dev sentences
 - Decode reordered test sentences

Better Translation

System	BLEU	TER
Moses + MSD	25.55	59.76
LOP Preprocessing + Moses	26.40	58.63
LOP Preprocessing + Moses + MSD	26.44	59.23

- 2000 German-English Europarl sentences
- Significant according to a paired permutation test

BLEU Improvement by Amount of Reordering



Conclusion

- Proposed a novel model of reordering for machine translation
 - Novel search and learning algorithms for this model
- Even as preprocessing, outperforms Moses' integrated reordering model
 - Could be integrated into many decoders
 - Could reorder phrases instead of words

Thanks!

- Much more in my dissertation:
 - More local search neighborhoods
 - Learning methods using likelihood/sampling
 - Methods for integrating LOP model into decoder
- <http://nlp.cs.jhu.edu/~royt/>