A Semi-supervised Word Alignment Algorithm with Partial Word Alignment

Qin Gao, Nguyen Bach and Stephan Vogel
Language Technologies Institute
Carnegie Mellon University

Supervised and unsupervised word alignment

Unsupervised word alignment:

- GIZA++ (IBM Models)
- LEAF (Fraser and Marcu, 2007)

Supervised word alignment

- Maximum Weight Bipartite Matching (Taskar et al 2005, Lacoste-Julien et al 2006)
- Maximum Entropy (Ittycheriah and Roukos 2005)

Semi-supervised word alignment

Discriminative models:

- Blunsom and
- Ni s and
- Tas

tive

)3)

006

Mix g parameters trained from labeled and unlabeled data

- Callison-Burch et al 2004

Manual alignments as references

Incomplete reference?

伊犁 大规模 开展 " 面对面 " 宣讲 活动

Xinjiang ‘s Yili holds propaganda drive

An oracle that only has partial knowledge
E.g. Location name dictionary, technical term dictionary

What is the difference of partial and full word alignment?

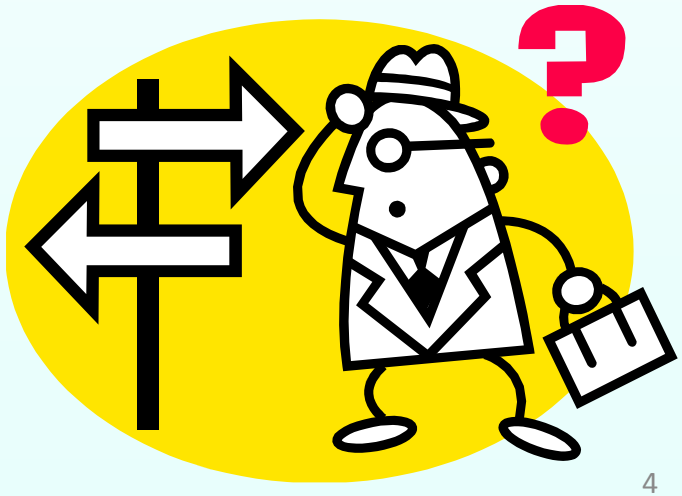
伊犁 大规模 开展 " 面对面 " 宣讲 活动

Xinjiang ‘s Yili holds propaganda drive

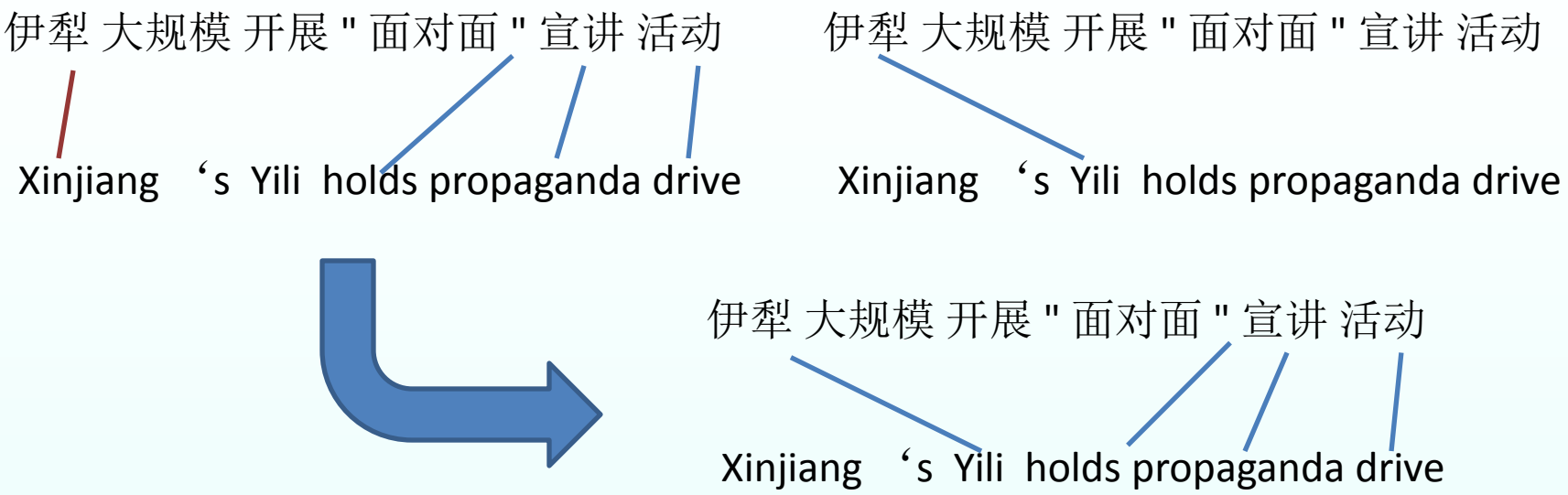
伊犁 大规模 开展 " 面对面 " 宣讲 活动

Xinjiang ‘s Yili holds propaganda drive

AER?
Entropy?
Likelihood?



How to make use of the partial knowledge?



Most straightforward way: Post-process

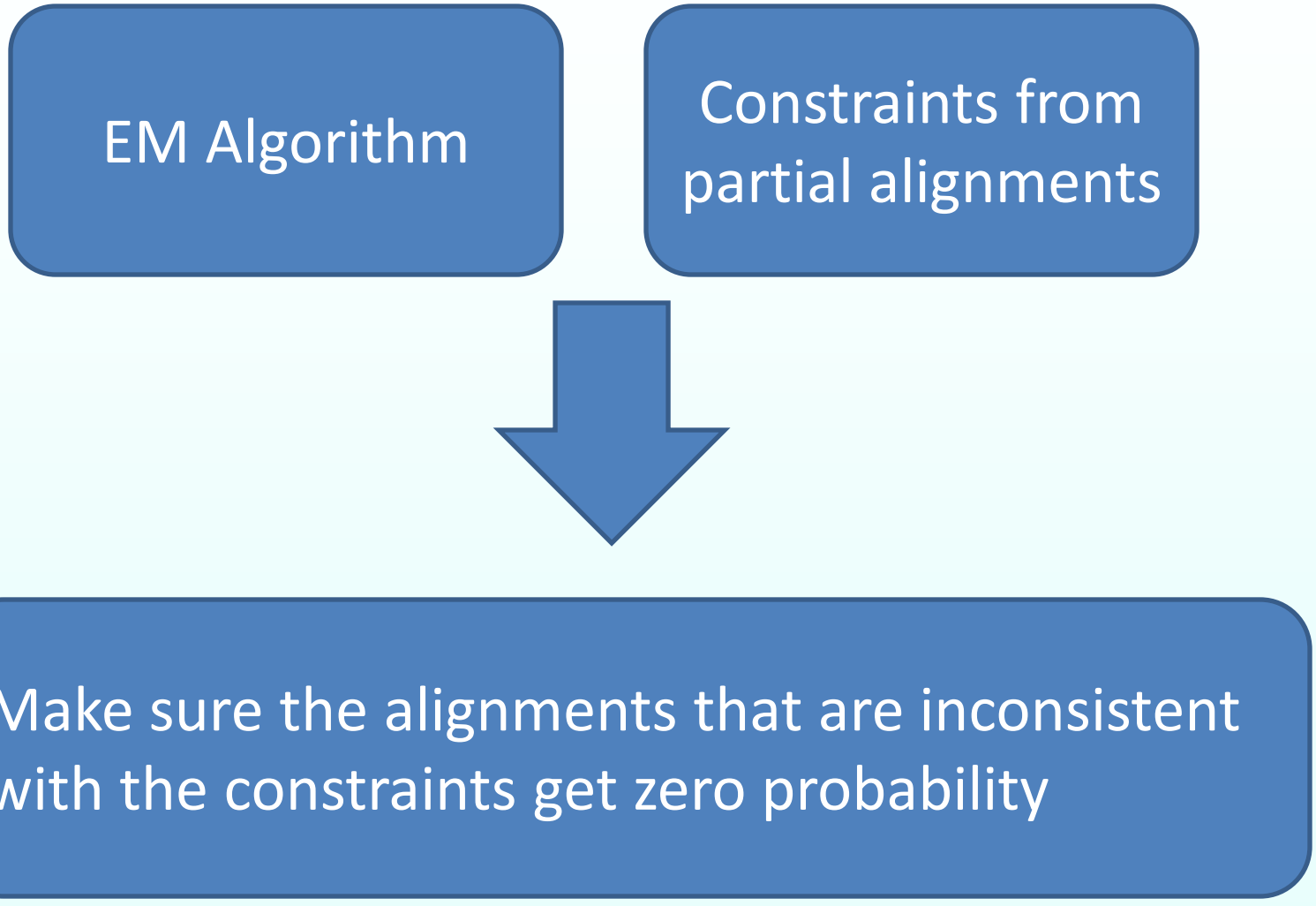
A better way

- Let the knowledge determine the known part, and let models determine the rest.
- The knowledge will:
  - Affect the statistics we get for the model
  - Be reflected in the final alignment

Anything conflicting with known alignments should be forbidden

- Pereira and Schabes, 1992, Similar idea on SCFGs

A summary of the proposed method

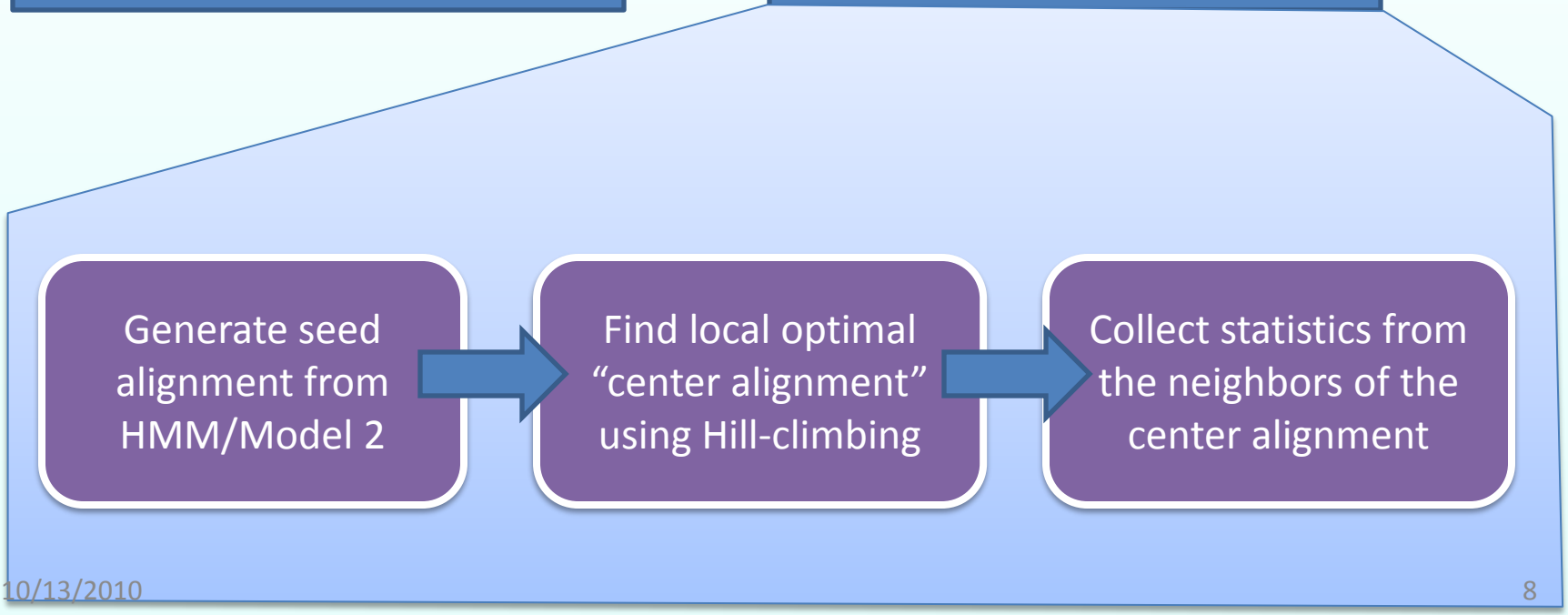


Recap: Training of IBM Models

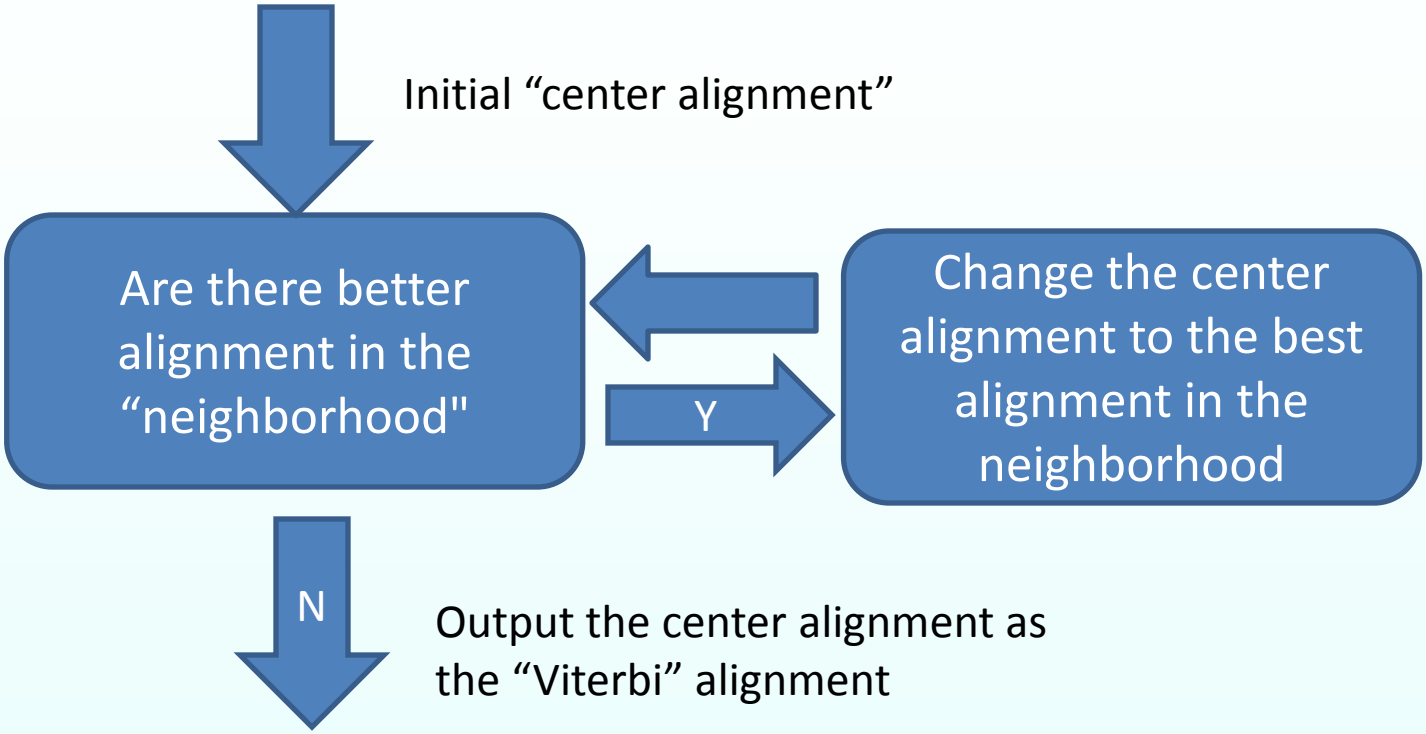
EM Algorithm

Model 1,2 :
Maximizing $L(F,A|E,\Theta)$ can be done directly over all possible alignments

Model 3,4,5 :
Even finding Viterbi alignment is too expensive



Hill-climbing



Neighborhood of an alignment

Move operator
$M_{1,1}$

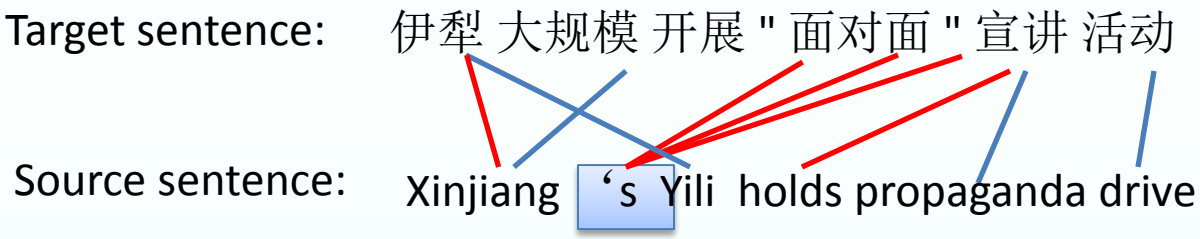
Swap operators
$S_{7,8}$

All the alignments that are different from the current center alignment by ***ONE*** operator, are considered in the neighborhood of it.

Applying the idea on fertility-based models

- Ruling out the inconsistent alignments during the hill-climbing stage
  - Yield a consistent, yet optimal “Viterbi” alignment
- Ruling out the statistics of inconsistent alignments during the statistics collection stage.
  - The probability of consistent alignments will increase in the next iteration

Two types of constraints



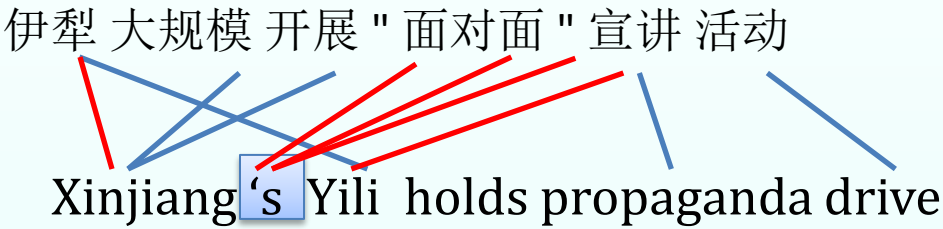
Type 1: A Target word aligned to a different source word

Type 2: A Source word is aligned to a concrete target word, but we know it should not be aligned to any.

Modified Hill-climbing

Every iteration, pick the alignment:
1. Corrects at least one inconsistent link
2. Has highest probability

Correct the inconsistent links



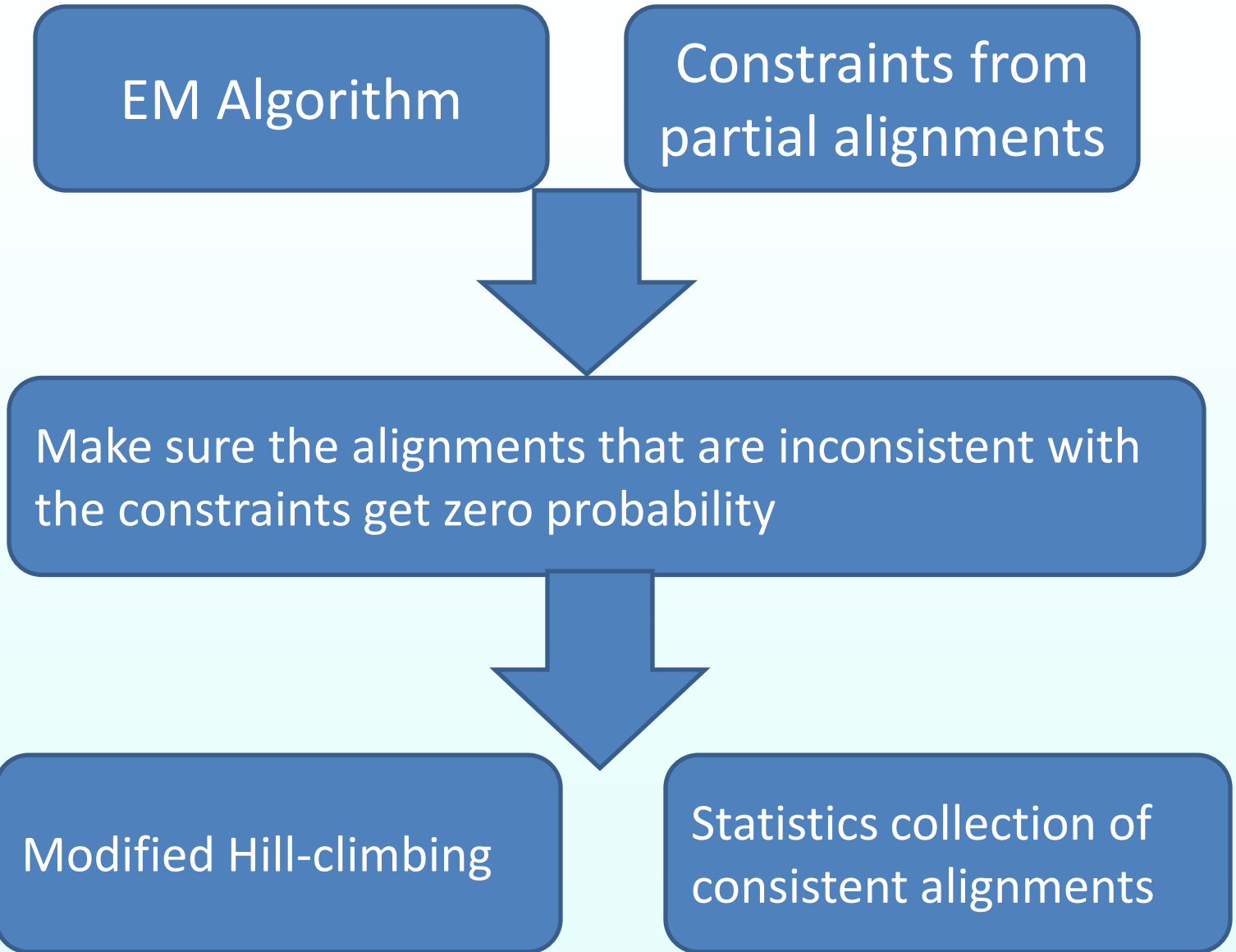
Optimize towards local optimal

Every iteration, pick the alignment:
1. Does not introduce inconsistent links
2. Has highest probability

Statistics collection

- Navigate through all the neighbor alignments and only collect statistics of consistent alignments.
- The modified hill-climbing and statistics collection can be done efficiently by manipulating the moving/swapping matrices

A summary of the proposed method



Experiment Design: What's the behavior of the method?

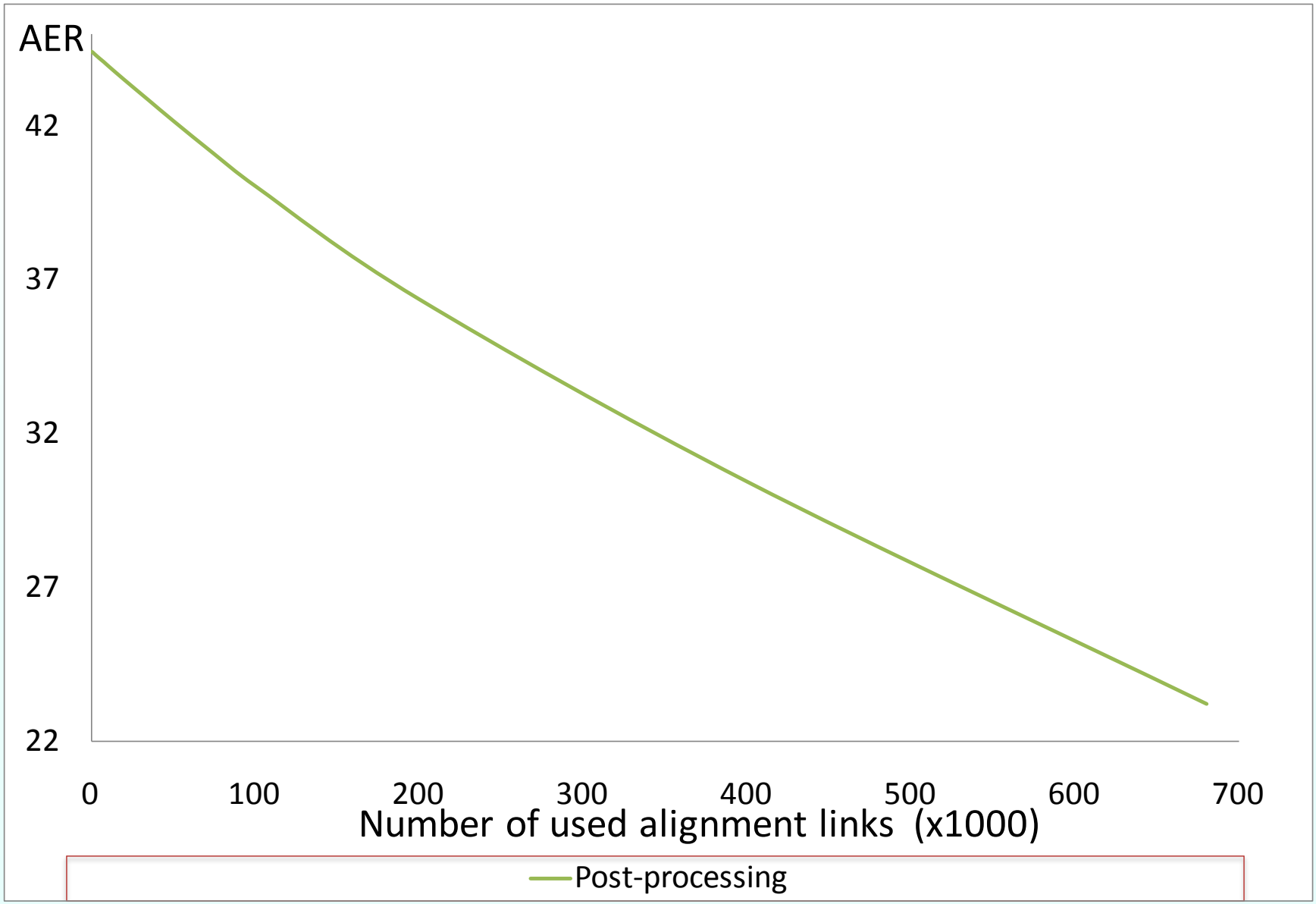
Does the algorithm correct more links than simple post processing?	→	We can use the method to exploit precious partial manual alignments.
Does having the answer for different part of the sentence makes difference?	→	We can use active learning methods to choose the most important words to label.
Is the method robust to the sub-optimal oracles (aka, error-prone answers)?	→	We can integrate the method with other automatic word aligners.

Experiment 1: does it correct more than what we already know?

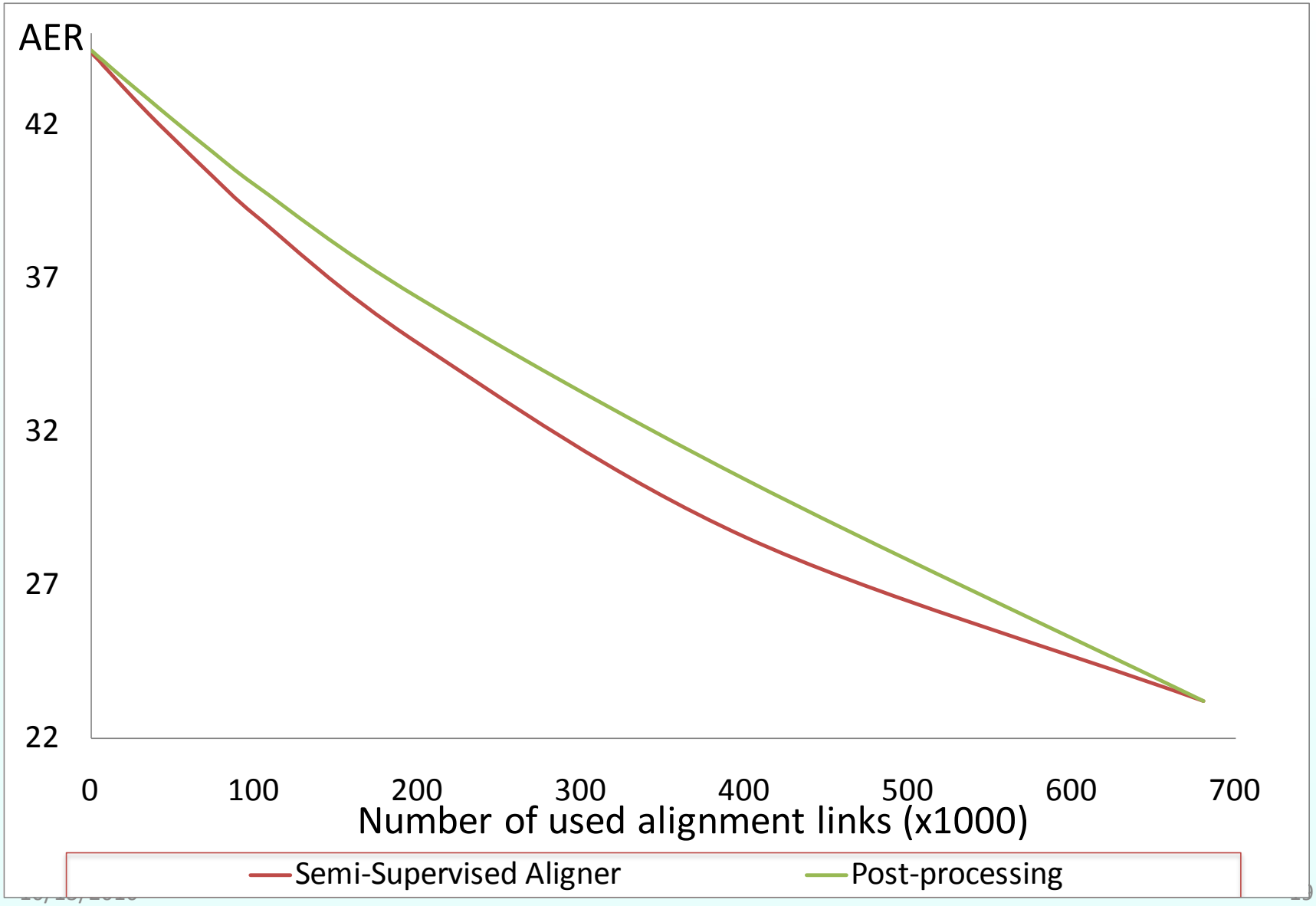
- The proposed method vs. post processing
- Randomly select different numbers alignment links
- Corpus: Chinese-English and Arabic-English manually aligned corpus

	Sentence Pairs	Num. of Words	Num. of Links
Chinese-English	21K	424K/524K	687K
Arabic-English	29K	630K/821K	830K

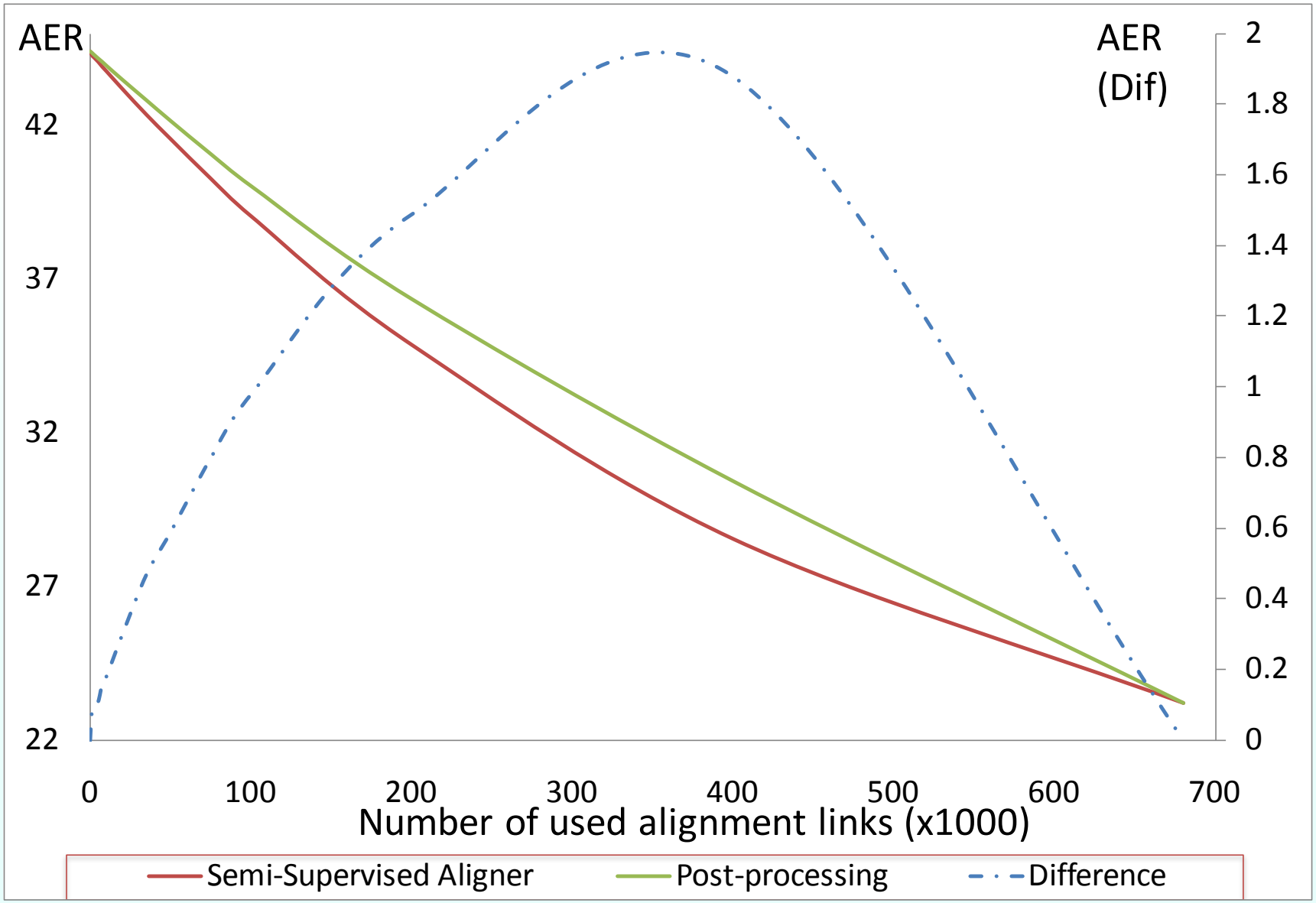
Alignment Error Rate Ch-En
Random Selection



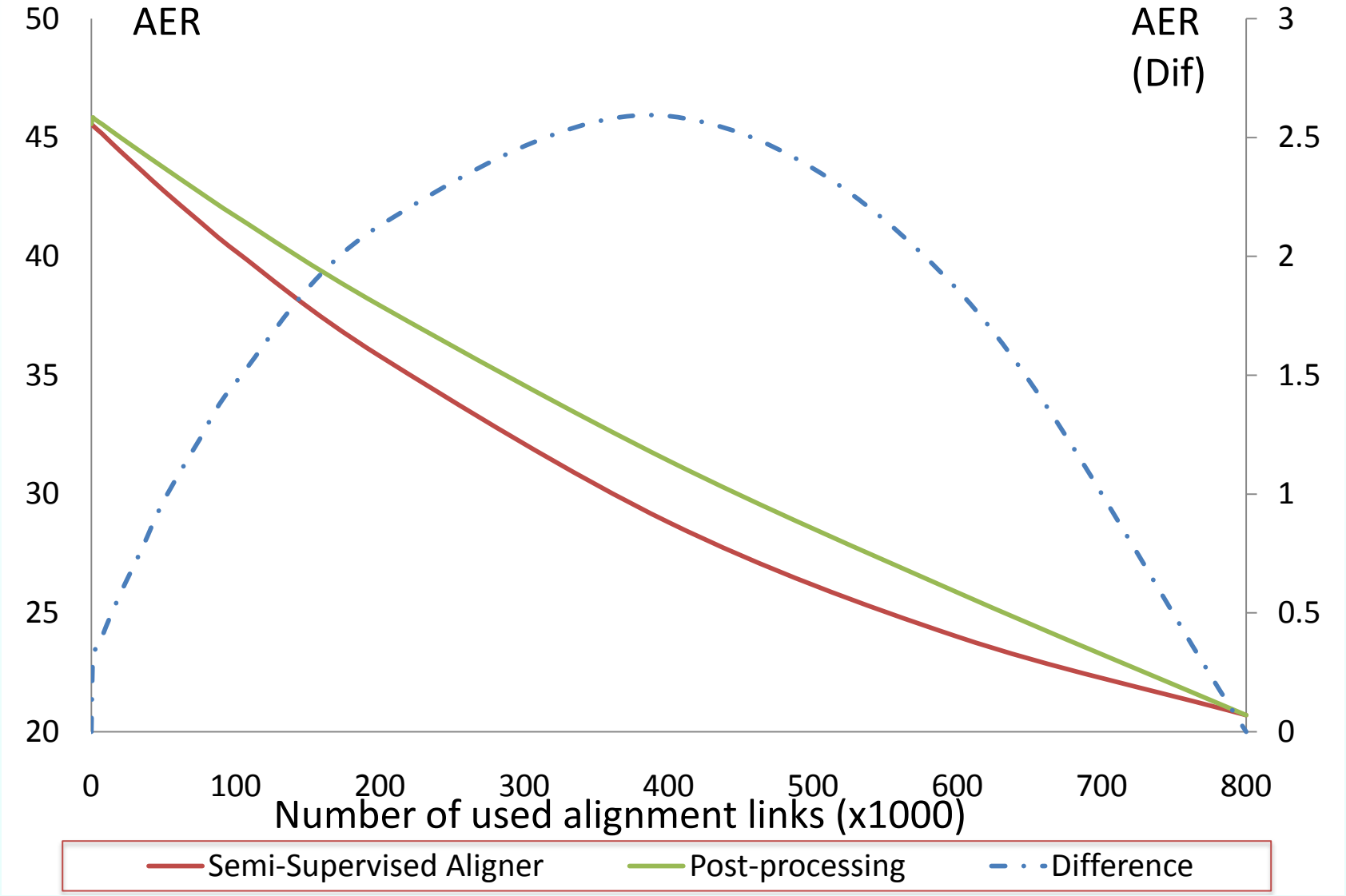
Alignment Error Rate Ch-En
Random Selection



Alignment Error Rate Ch-En
Random Selection



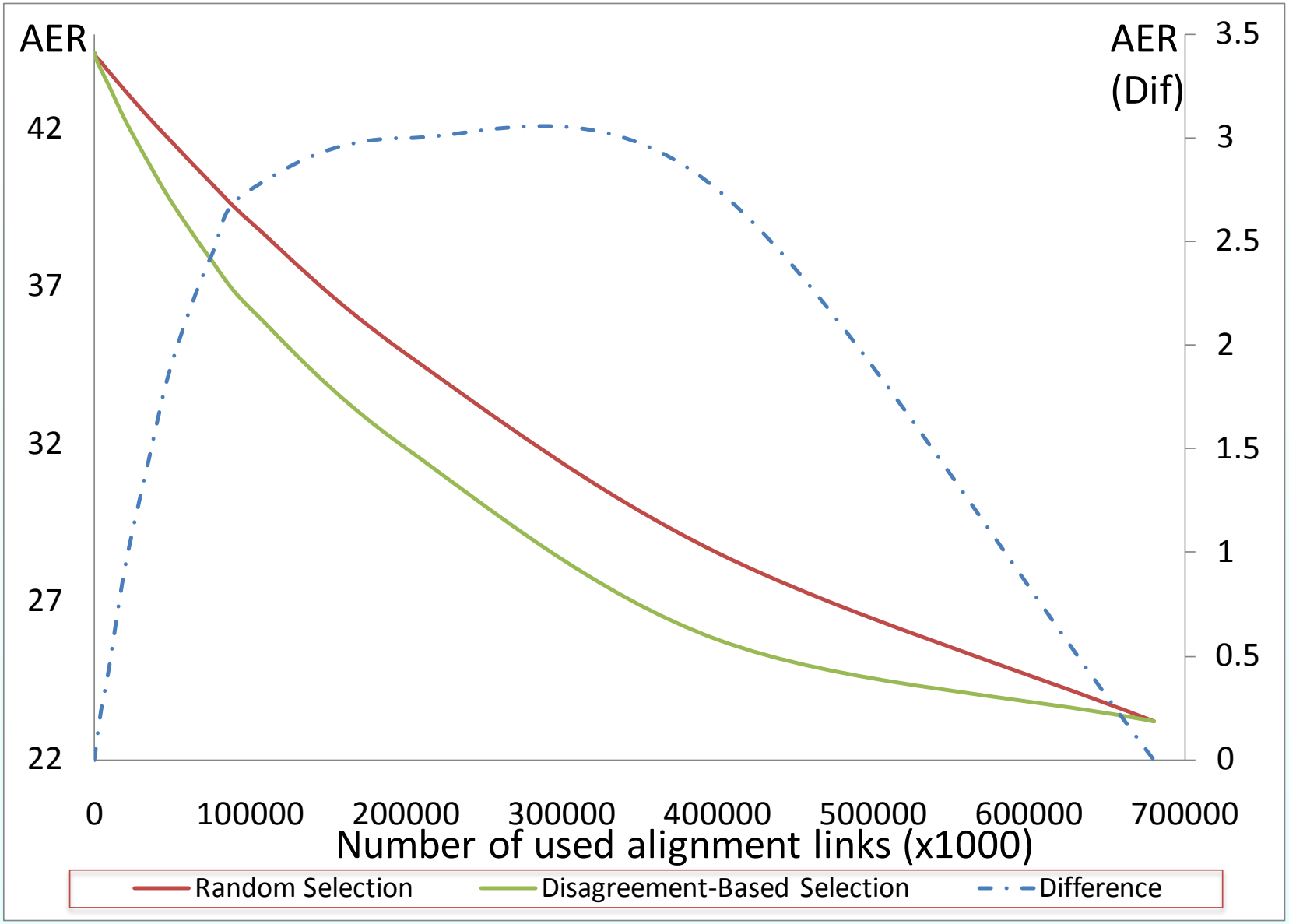
Alignment Error Rate Ar-En
Random Selection



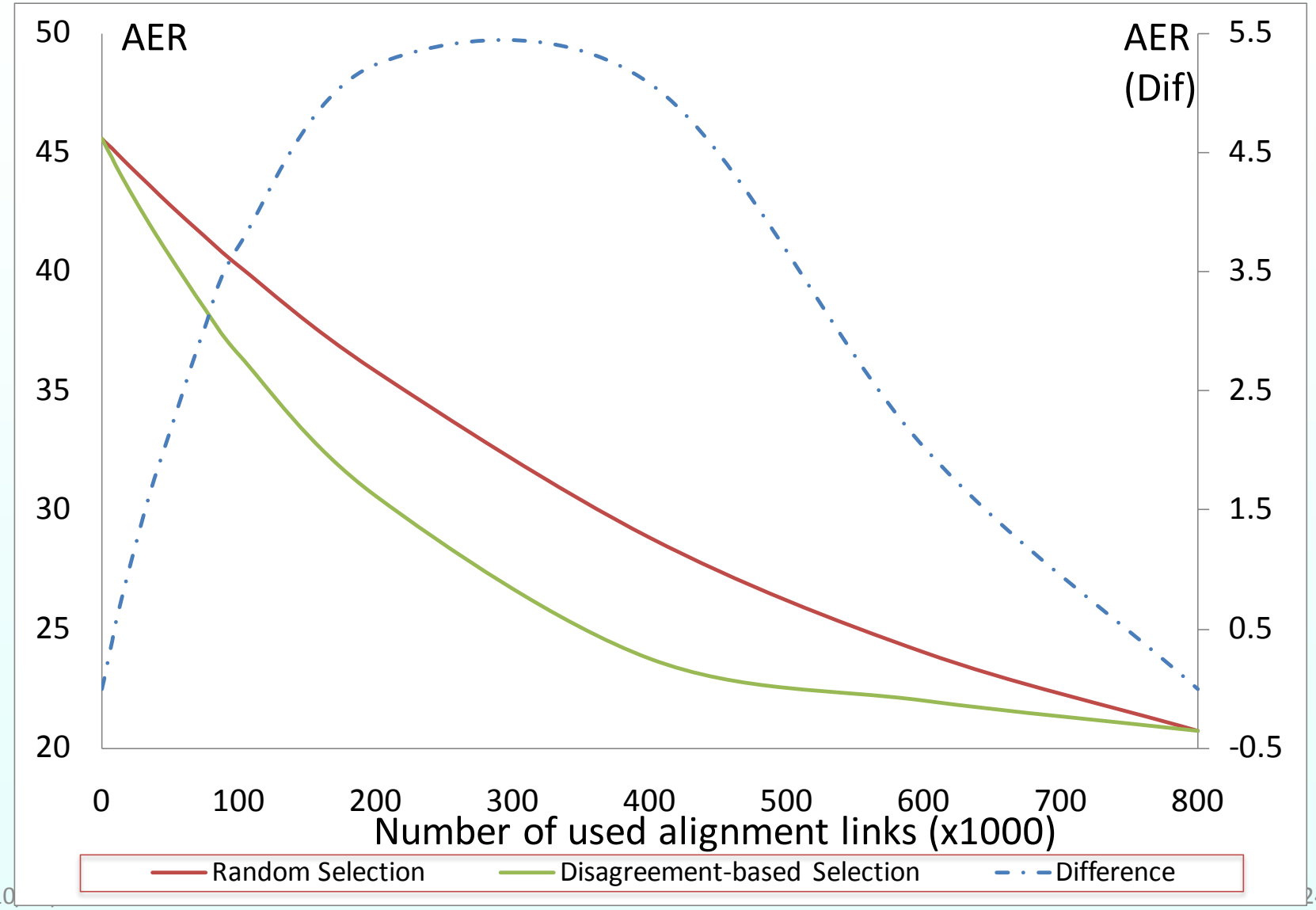
Experiment 2: Carefully selecting alignment links makes a difference

- We try to find a better way of selecting alignment links from manual alignment
- Intuitively: Select alignment links that two GIZA++ directions disagree about.

Alignment Error Rate Ch-En
Random Selection v.s Disagreement-based



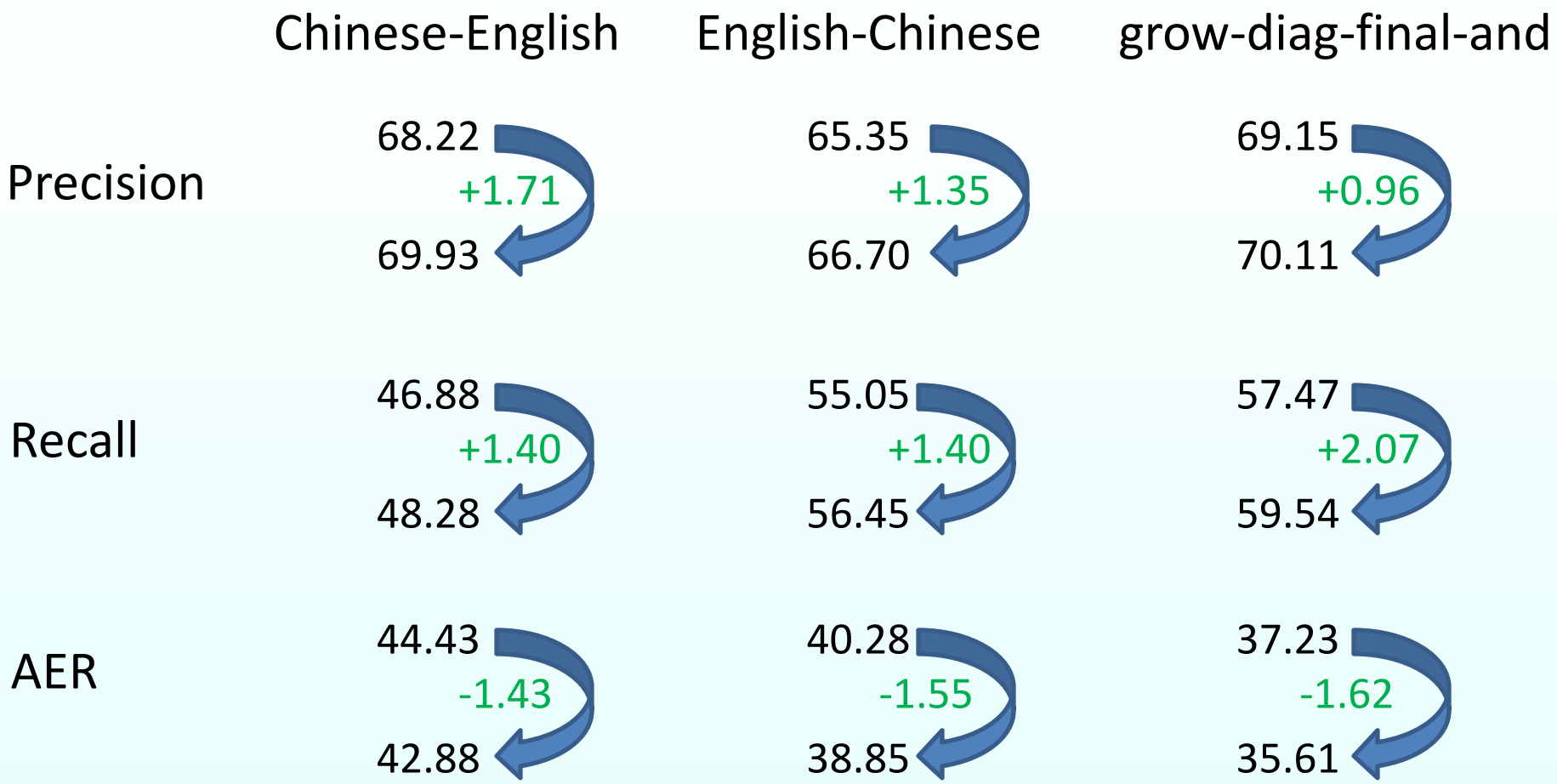
Alignment Error Rate Ar-En
Random Selection v.s Disagreement-based



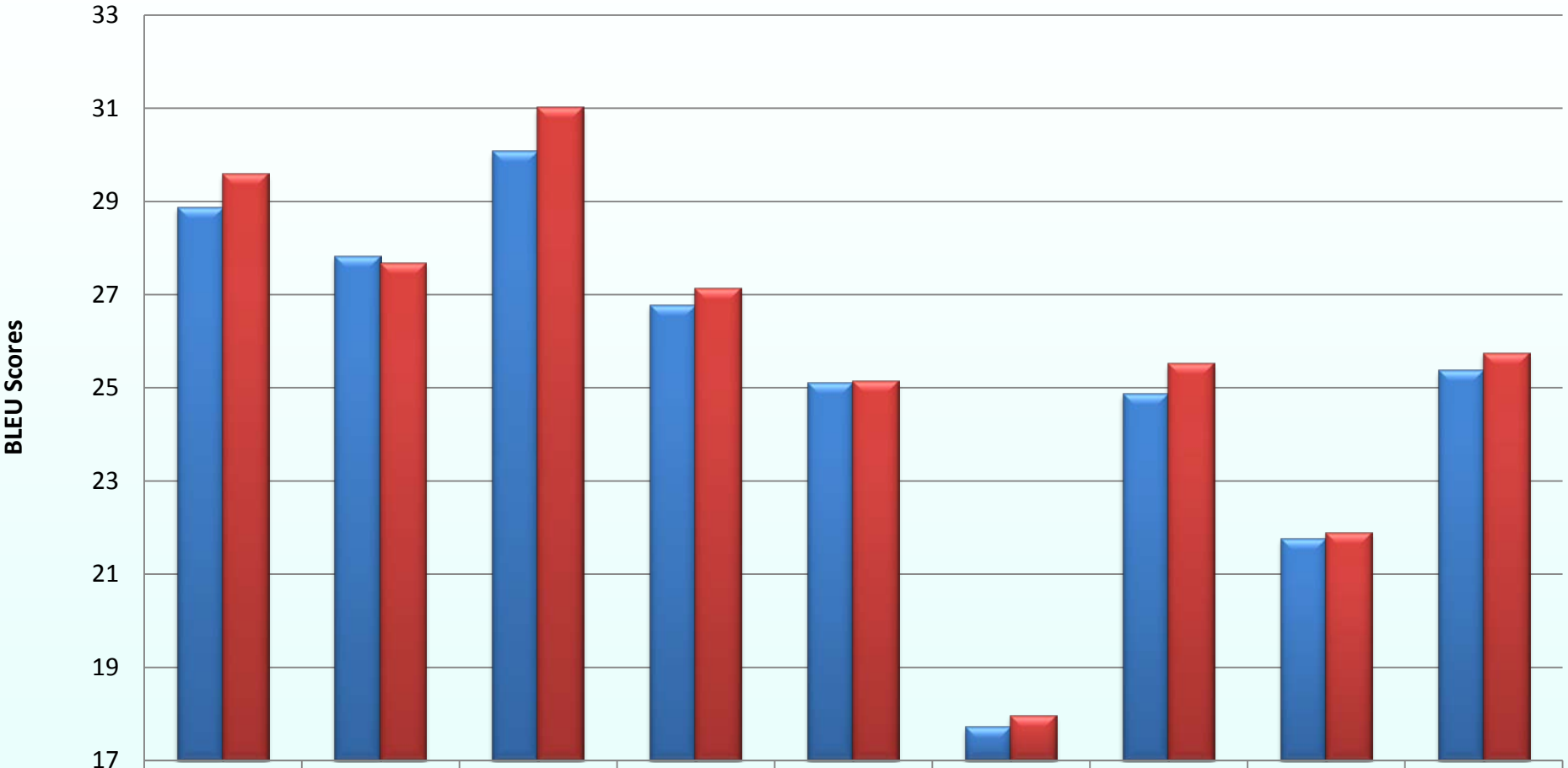
Experiment 3: Does it works for simple (AKA error-prone) heuristics?

- We used LDC Chinese-English dictionary to generate high-precision-low-recall partial alignments
- The entries with single Chinese character or more than six English words are filtered out.
- Add links when a lexicon entry was encountered in the sentence pair
- 79.48% precision and 17.36% recall rate

Alignment quality



The improvement on translation qualities (400k sentence pairs)



	MT02	MT03	MT04	MT05	MT08NW	MT08WB	Dev07NW	Dev07WB	Average
Baseline	28.87	27.82	30.08	26.77	25.09	17.72	24.88	21.76	25.37
Dict-Link	29.59	27.67	31.01	27.13	25.14	17.96	25.51	21.88	25.74

Conclusion

- We implemented a semi-supervised word alignment algorithm based on IBM models which can use partial word alignment.
- Experiments were performed to prove that:
  - 1. The algorithm can correct more links than directly fixing the incorrect links
  - 2. Better alignment quality can be achieved by carefully selecting words to ask the oracle
  - 3. By supplying high-precision-low-recall alignment links the alignment quality can also be improved.

Applications

- Integrate manual alignments
  - Use Mechanical Turk to collect partial alignments
  - Reference: Gao and Vogel, MTurk workshop 2010
- Active learning for word alignment
  - Select the most significant words to ask the oracles for answers
  - Reference: Ambati, Vogel and Carbonell, ACL 2010
- Interact with supervised aligners
  - Use supervised aligner to generate high-precision-low-recall alignment links
  - Iteratively improve both aligners
  - Reference: Gao, Guzman and Vogel, COLING 2010

THANK YOU

The improvement on alignment qualities

Chinese-English			
	Precision	Recall	AER
Baseline	68.22	46.88	44.43
Dict-Link	69.93	48.28	42.88
English-Chinese			
	Precision	Recall	AER
Baseline	65.35	55.05	40.24
Dict-Link	66.70	56.45	38.85
grow-diag-final-and			
	Precision	Recall	AER
Baseline	69.15	57.47	37.23
Dict-Link	70.11	59.54	35.61

Priority of alignment links

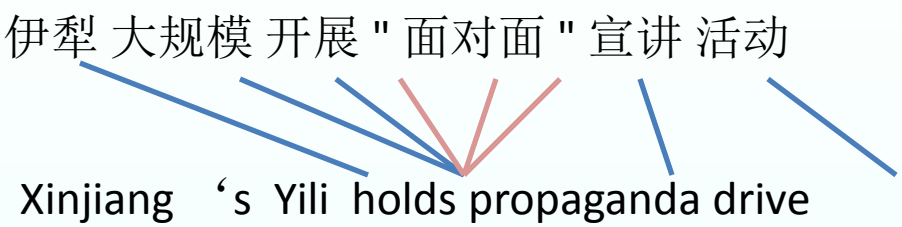
- Run GIZA++ on both directions
- Pick the words that have different alignments in two directions and ask for answers from the oracle

- $C1$: f_j aligns to $e_i, i > 0$ in $e \rightarrow f$,[1] but in reversed direction e_i does not align to f_j but to another word.
- $C2$: f_j aligns to $e_i, i > 0$, in $f \rightarrow e$, but in reversed direction $(e \rightarrow f)$, f_j aligns to the empty word.
- $C3$: no word aligns to f_j, in $f \rightarrow e$, but in reversed direction f_j aligns to $e_i, i > 0$.[2]

Order	Criterion	Order	Criterion
1	$f_j \in C1$	5	$e_i \in C2$
2	$f_j \in C2$	4	$e_i \in C1$
3	$f_j \in C3$	6	$e_i \in C3$

Table 1: The priorities of alignment links

The result by applying our method



Fix the known alignment links during training, instead of fixing it afterwards.

By fixing the link between Yili and伊犁, other links also get corrected. (Of course, new errors may be introduced)

Applications

- Using active learning method to collect useful alignment links from users
- Make use of noisy manual alignments from non-experts such as Mechanical Turk
- Integrate other word aligners e.g. discriminative word aligner

In the context of word alignment

$$P(f_1^J|e_1^I) = \sum_{a_1^J} Pr(f_1^J, a_1^J|e_1^J, \theta)$$

$$Pr(f_1^J|e_1^I, a_1^J, \alpha_I^J) = \begin{cases} 0, a_1^J \text{is inconsistent with} \alpha_I^J \\ Pr(f_1^J|e_1^I, a_1^J, \theta), \text{otherwise} \end{cases}$$

Overview

- A word alignment algorithm that can take partial alignment links and improve the unsupervised word alignment results
- The method enables integrating knowledge from external sources into unsupervised word alignment

In this talk

- The detail of the algorithm
- Integrating (oracle) manual alignments
- Applying alignment links from lexicon

Easy things easy (corrected)
Hard things possible (improved)

What is partial word alignment?

Modified Hill Climbing

- In Hill-climbing step 1 : Pick the operator that
  1. eliminate at least one inconsistent link
  2. yield best probability

Modified Hill Climbing

- In Hill-climbing step 2:
  - Make a cell negative if the operator will bring in inconsistent link
  - Recalculate the signs of the cells in swap matrix if a swap operator is chosen

The M-Step

- In statistics collection :
  - Ignore the statistics from cells with negative values