Three models for discriminative machine translation using Global Lexical Selection and Sentence Reconstruction

Sriram Venkatapathy

IIIT – Hyderabad

Srinivas Bangalore

AT & T Research Labs

Complexity of the task

► People of these islands have adopted Hindi as a means of communication .

► इन द्वीपों के   लोगों ने हिंदी भाषा को एक संपर्क भाषा के रूप में अपना लिया है .

► These islands of people   hindi language   a commu. language in form of   adopted-take-be

► Primary Observation:

  ▪ There are long distance word order variations in English-Hindi unlike English-French.

2

Outline

► **Previous Work**

► Global Lexical Selection

► Three models

  ▪ Bag-of-Words Lexical Choice Model
  ▪ Sequential Lexical Choice Model
  ▪ Hierarchical Lexical association and Reordering Model

► Results

► Conclusion and Future Work

Previous work on Stat MT.

► Local associations between source and target phrases are obtained.

1. GIZA++ is used to align source words to target words.

2. These alignments augmented with target-to-source alignments.

3. Word-alignments are extended to obtain phrase level local associations.
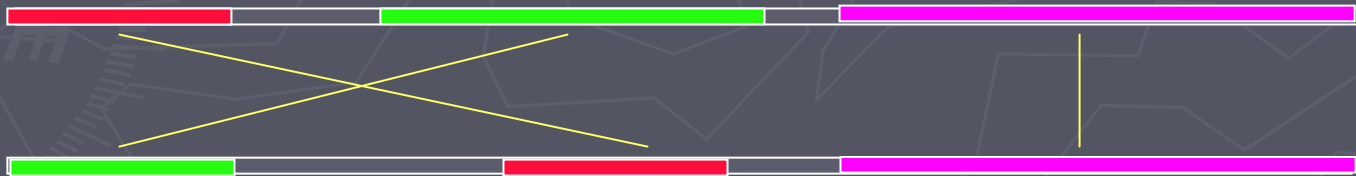
Previous work on Stat MT.

► Translation is done in two steps

1. Local associations of phrases of source sentence are selected.
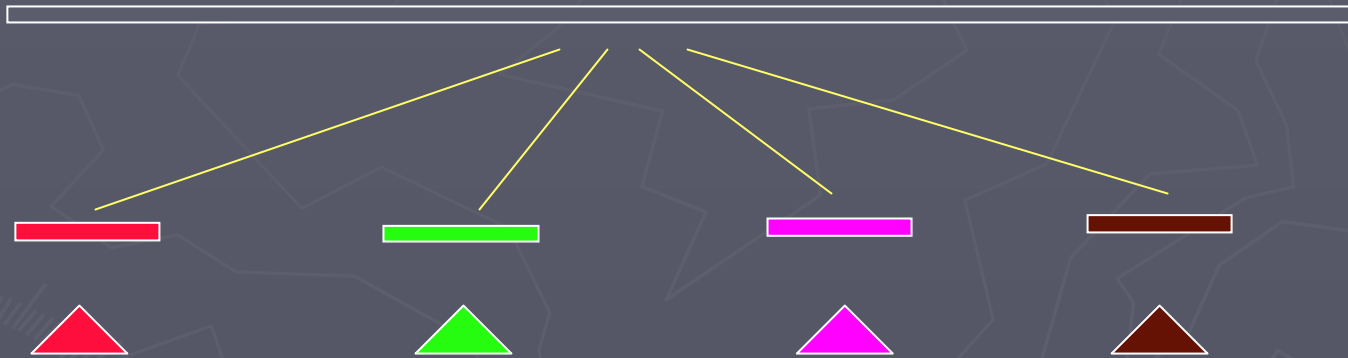
2. Re-ordering the target language phrases.

Outline

► Previous Work

► Global Lexical Selection

► Three models

  ▪ Bag-of-Words Lexical Choice Model
  ▪ Sequential Lexical Choice Model
  ▪ Hierarchical Lexical association and Reordering Model

► Results

► Conclusion and Future Work

Global Lexical Selection

► In contrast, the target words are associated to the entire source sentence.

► Intutions

1. Lexico-syntactic features (not necessarily single words) in source sentence might trigger the presence of target words.

2. Also predict syntactic cues along with lexical/phrasal units.

Global Lexical Selection

► No longer tight association between source language words/phrases.

► During translation,
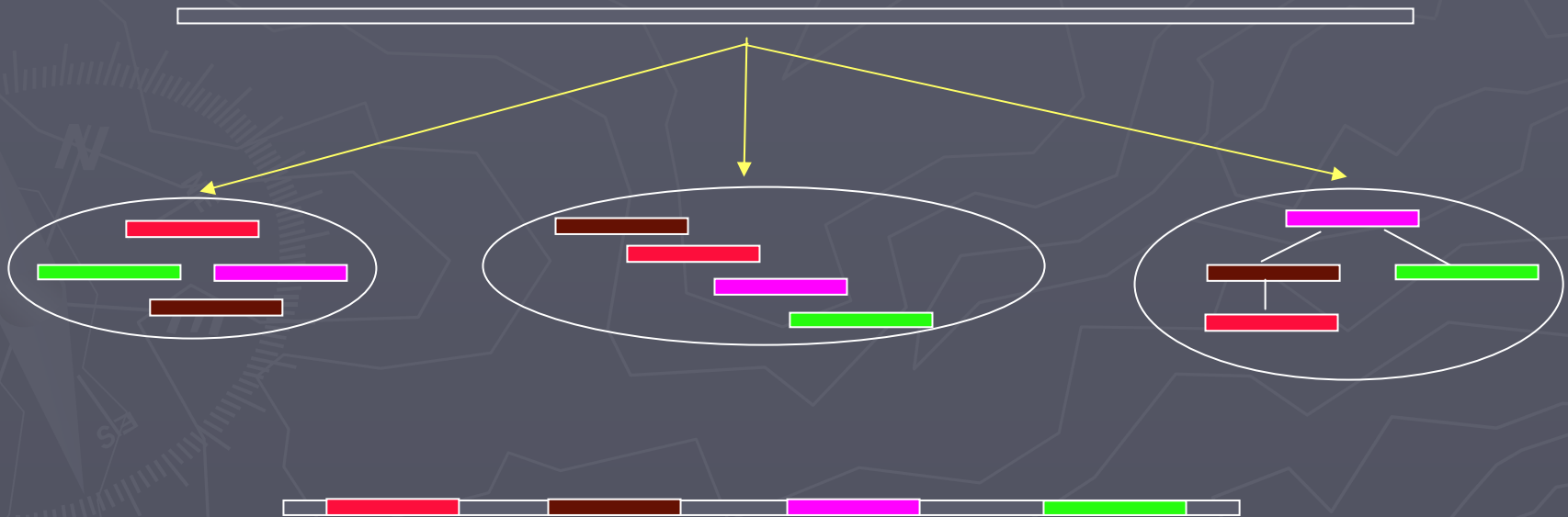
Outline

▶ Previous Work

▶ Global Lexical Selection

▶ Three models

- ▪ Bag-of-Words Lexical Choice Model
- ▪ Sequential Lexical Choice Model
- ▪ Hierarchical Lexical association and Reordering Model

▶ Results

▶ Conclusion and Future Work

Bag of words model

► Learn: Given a source sentence S, what is the probability that a target word t is in its translation ?

   i.e., estimate $p(\text{true} \mid t, S)$ and $p(\text{false} \mid t, S)$

► Binary classifiers are built for all words in target language vocabulary.

► Maximum entropy model is used for learning.

Bag of words model - Training

► Training binary classifier for target language word   t.

► Example sentences:

s1	──────────────	True (t exists in translation)
s2	──────────────	False (t doesn't exists in translation)
s3	──────────────	False (t doesn't exists in translation)
s4	──────────────	True (t exists in translation)

► Number of training sentences for each target language word are total number of sentence pairs.

Bag of words model – Lexical selection

► For an input sentence S, first the target sentence bag is obtained.

► Source sentence features considered : N-grams

   ▪ Let, BOgrams(S) be N-grams of source sentence S.

► The bag contains a target word w, if

$$p \text{ (true | t, BOgrams(S))} > \tau \text{ (threshold)}$$

► BOW (T) = { t | p (true | t, S) > τ)

Bag of words model
– Sentence Reconstruction

► Various permutations of words in BOW (T) considered and then ranked by a target language model.

► All possible permutations -- computationally not feasible.

► Reduced by constraining permutations to be within local window of adjustable size (perm) .   (Kanthak et al., 2005)

► During decoding, some words can be deleted. Parameter (δ) can be used to adjust length of translated outputs.

Outline
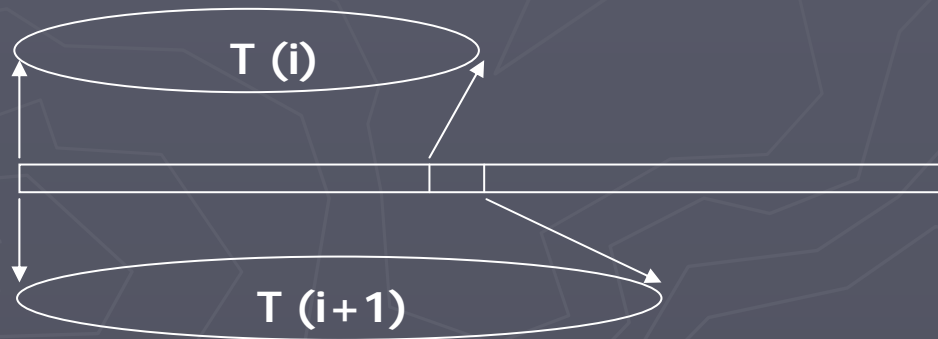
► Previous Work

► Global Lexical Selection

► Three models

  ▪ Bag-of-Words Lexical Choice Model
  ▪ Sequential Lexical Choice Model
  ▪ Hierarchical Lexical association and Reordering Model

► Results

► Conclusion and Future Work

Sequential lexical choice model

► In Previous approach we begin permuting with an arbitrary order of words as start point.

► Better to start with a more definite string.

► During lexical selection, target words are first placed in an order faithful to source sentence words.

► Training same as bag of words model.

Sequential model - Decoding

► Goal: Associate sets of target words with every position in source sentence (S).

► Predict bags of words (T_i) for all prefixes of S.

T (i)

T (i+1)

► Associate a target word t to source position (i+1) if it is present in T_{i+1} but not in T_i .

Sequential model - Decoding

► Intution:

  ▪ Word t associated with position i if some information at ith position triggered it.

► Example:
  ► Pay a   :   दो
  ► Pay a visit   :   मिलो

  ▪ Associate मिलो with the position of **visit** in source sentence.

► **Limitation**
 ▪ **Using moving permutation window can explore only local word reordering.**

Outline

► Previous Work

► Global Lexical Selection

► Three models

  ▪ Bag-of-Words Lexical Choice Model
  ▪ Sequential Lexical Choice Model
  ▪ Hierarchical Lexical association and Reordering Model

► Results

► Conclusion and Future Work

Hierarchical model

► Sequential model expected to work better for language pairs with only local word order variations.

  ▪ May perform poorly for language pairs (English-Hindi) with significant word order variation.

► Previous approach: Associated target words with source positions.

► This approach: Associate target words with nodes of source dependency tree.

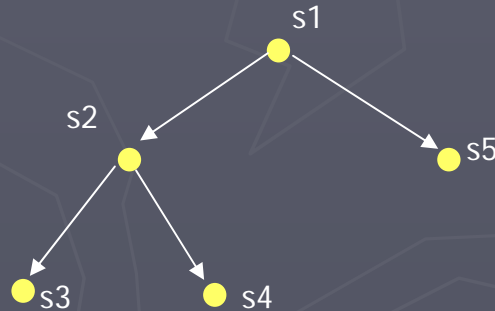# Hierarchical model - attachment
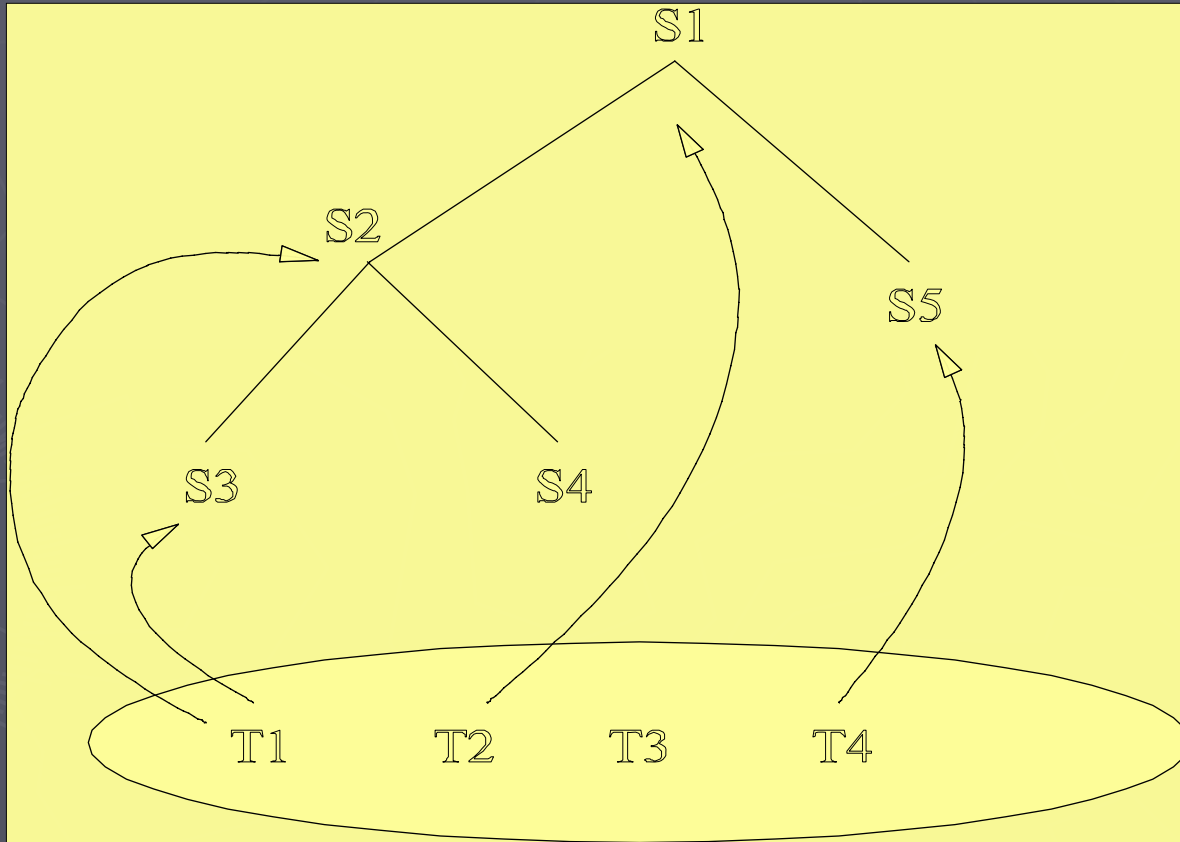
Hierarchical model - decoding

1. Predict the bag-of-words (same as previous models)

   ► Given source sentence S and its dependency structure.

2. Attachment to source nodes.

   ► Attach words from previous step to various nodes of source dependency structure.

3. Ordering target language words

   ► Traverse source dependency structure in bottom-up fashion to obtain best target string.

Predict Bag-of-words

► Same as bag-of-words model except that both n-gram features and dependency features are used.

▪ Include t   if ,   p (true| t, f (S)) > τ

► Features f(S)

s1

s2
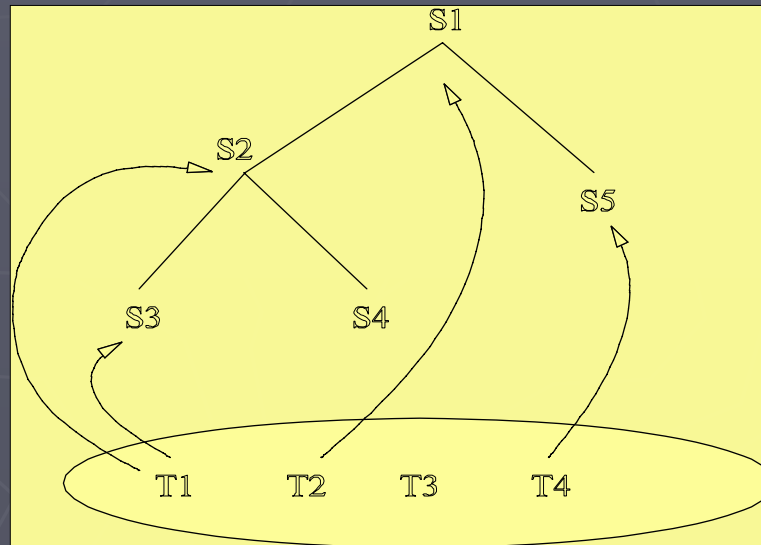
s5

s3   s4

▪ N-gram features   's1'   's2'   's3 s2'   's2 s4 s1'

▪ Dependency pairs   's2 s1'   's4 s2'

▪ Dependency Treelet   's3 s2 s5'   's2 s1 s5'

22

Hierarchical model - attachment

► Every target word t is attached to the source node whose *local features* give the best positive probability for word t.

► If s_t is source node to which target word t is attached to

▪ s_t = argmax $_s$ p(true| t, f_L(s)) .

Hierarchical model – ordering

► Source sentence dependency tree is traversed in a bottom-up fashion.

► The best target string for every sub-tree is determined.
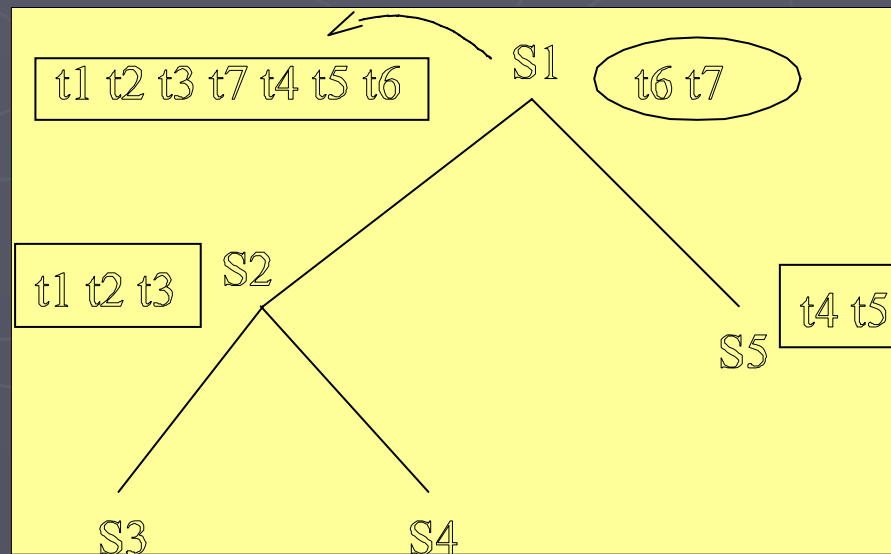
Outline

► Previous Work

► Global Lexical Selection

► Three models

   ▪ Bag-of-Words Lexical Choice Model
   ▪ Sequential Lexical Choice Model
   ▪ Hierarchical Lexical association and Reordering Model

► Results

► Conclusion and Future Work

Experiments - Dataset

- ➤ Language pair: English – Hindi (large word-order variations)

- ➤ Training set : 37967 pairs
- ➤ Development : 819 pairs
- ➤ Test : 699 pairs

- ➤ Maximum sentence length = 30

- ➤ Unseen tokens in target side of devel corpus : 13.48%
- ➤ Unseen tokens in source side of devel corpus : 10.77%
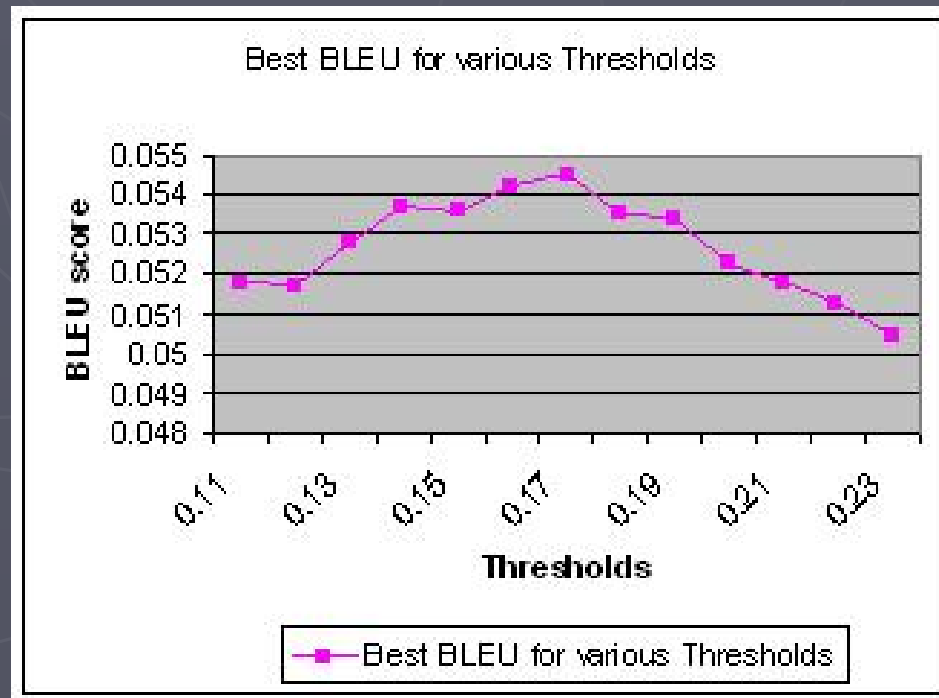
Results – Bag of words model

► Need to determine the best value of т, perm and δ .

► Quality of bags (Lexical accuracy/F-score) determined by threshold т.   Best LexAcc = 0.455.

Lexical Accuracy of Predicted Bags of Words
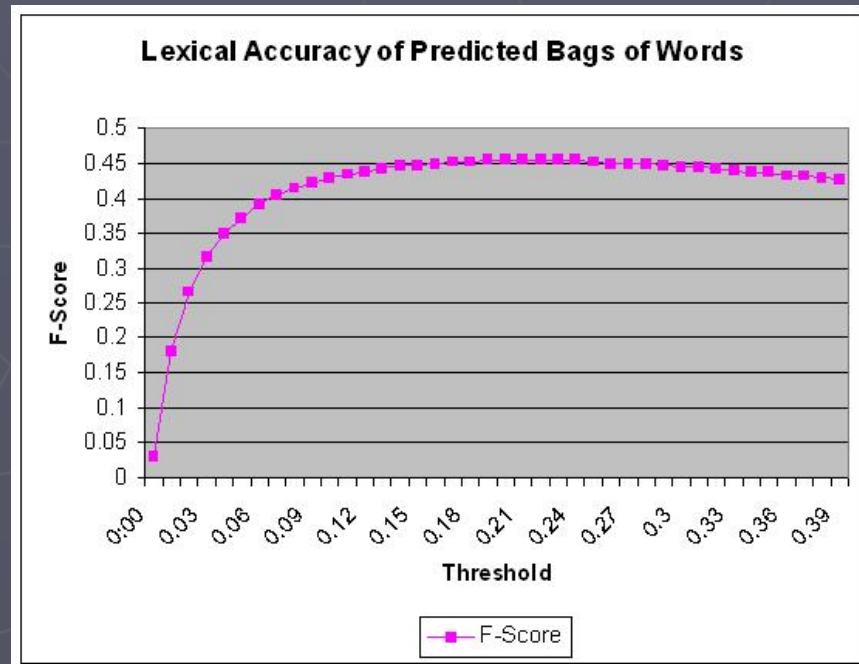
Results - Bag of words model

► All the bags obtained using various thresholds are now permuted.

► Best Bleu scores for various thresholds. Best Bleu = 0.0545

Results

		Devel. Set	Test. Set	
		BLEU	LexAcc	BLEU
Bag of words		0.0545	46.20	0.0428
Sequential		0.0586	45.24	0.0473
Hierarchical		0.0650	46.20	0.0498
MOSES (3 → 1)			34.42	0.0381
MOSES (3 → 3)			32.18	0.0440
MOSES (7 → 7) (Untuned)			28.23	0.0222

Conclusion

► Global Lexical selection

- To Make use of lexico-syntactic features on source.
- Predict syntactic cues along with lexical/phrasal units.

► Predicted units are semi-aligned with source structures for better target sentence re-construction.

- Alignment is an inferred step and not a primary step.

► These models give scope for obtaining entirely different structures in target language.

Future work

- ► Improve hierarchical reordering model.
  - ▪ Take K-best target strings for every sub-tree during traversal.

- ► Handling cases of structural non-isomorphism between source and target sentences

- ► Consider phrases on target sentence instead of just words.

Three models for discriminative machine translation using Global Lexical Selection and Sentence Reconstruction

Sriram Venkatapathy

IIIT – Hyderabad

Srinivas Bangalore

AT & T Research Labs