

Discriminative word alignment by learning the alignment structure and syntactic divergence between a language pair

Sriram Venkatapathy
IIIT – Hyderabad

Aravind Joshi
University of Pennsylvania

Outline

- ▶ Word Alignment - English-Hindi Language Pair
- ▶ Related approaches
- ▶ Discriminative Re-ranking approach
 - Features
 - Parameter optimization using MIRA
 - Results
- ▶ Future Work and Conclusion

Word-Alignment

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

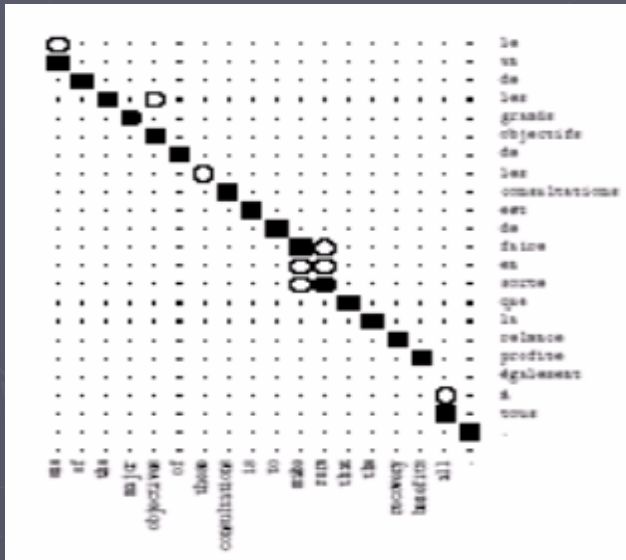
- ▶ इन द्वीपों के लोगों ने हिंदी भाषा को एक संपर्क भाषा के रूप में अपना लिया है .

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

- ▶ Primary Observation:

- The alignment between English-Hindi is largely non-monotonic, unlike the alignment between English-French.

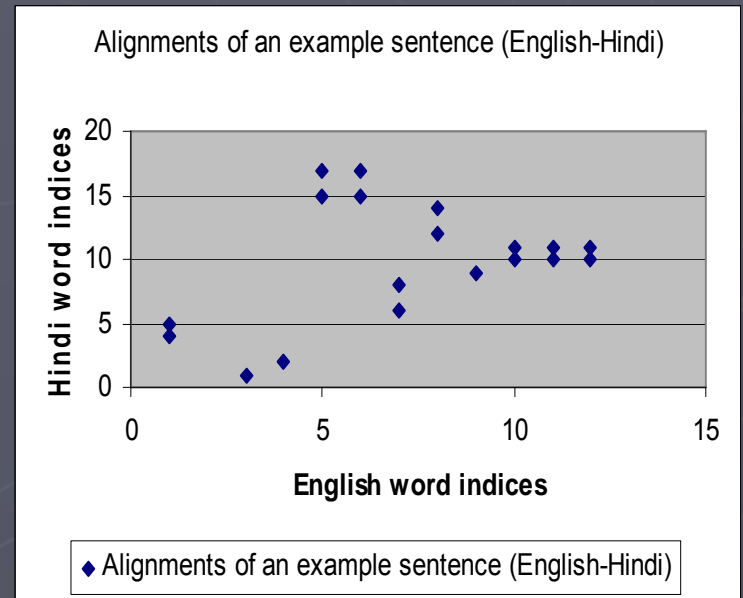
Comparison



French

Hindi

English



English

Outline

- ▶ Word Alignment - English-Hindi Language Pair
- ▶ Related approaches
- ▶ Discriminative Re-ranking approach
 - Features
 - Parameter optimization using MIRA
 - Results
- ▶ Future Work and Conclusion

Related approaches

▶ Generative models

- IBM Models, HMM models (Implemented in Giza++)

▶ Discriminative models

- (Taskar et al., 2005)
- (Moore et al., 2005)

Generative models - Limitations

- Difficult to add new Parameters.
 - ▶ The generative story needs to be modified appropriately to incorporate the new parameters.
- Parameters are not optimized.
 - ▶ All the parameters used have equal weights. For example, translation probabilities have the same importance as distortion probabilities.
 - ▶ As more complex features are added to the model, the parameters need to be optimized appropriately.

(Taskar et al., 2005) - Limitations

- The alignment search and optimization requires that the features are local to the alignment link.
- There is 0th order correlation with other alignments links in an alignment.
- (Lacoste-Simon et al., 2006) include first-order features (similar to HMM Parameter) and fertility but still there isn't much room for more complex global features required for aligning diverse language pairs such as English-Hindi.

(Moore et al., 2006) - Limitations

- ▶ Structural features are applied on partial structures (ie., every time a new alignment link is considered)
 - May lead to ruling out good alignments at an early stage.
 - Restricts us from using more complex syntactic features. (As it is a left to right search).

Outline

- ▶ Word Alignment - English-Hindi Language Pair
- ▶ Related approaches
- ▶ Discriminative Re-ranking approach
 - Features
 - Parameter optimization using MIRA
 - Results
- ▶ Future Work and Conclusion

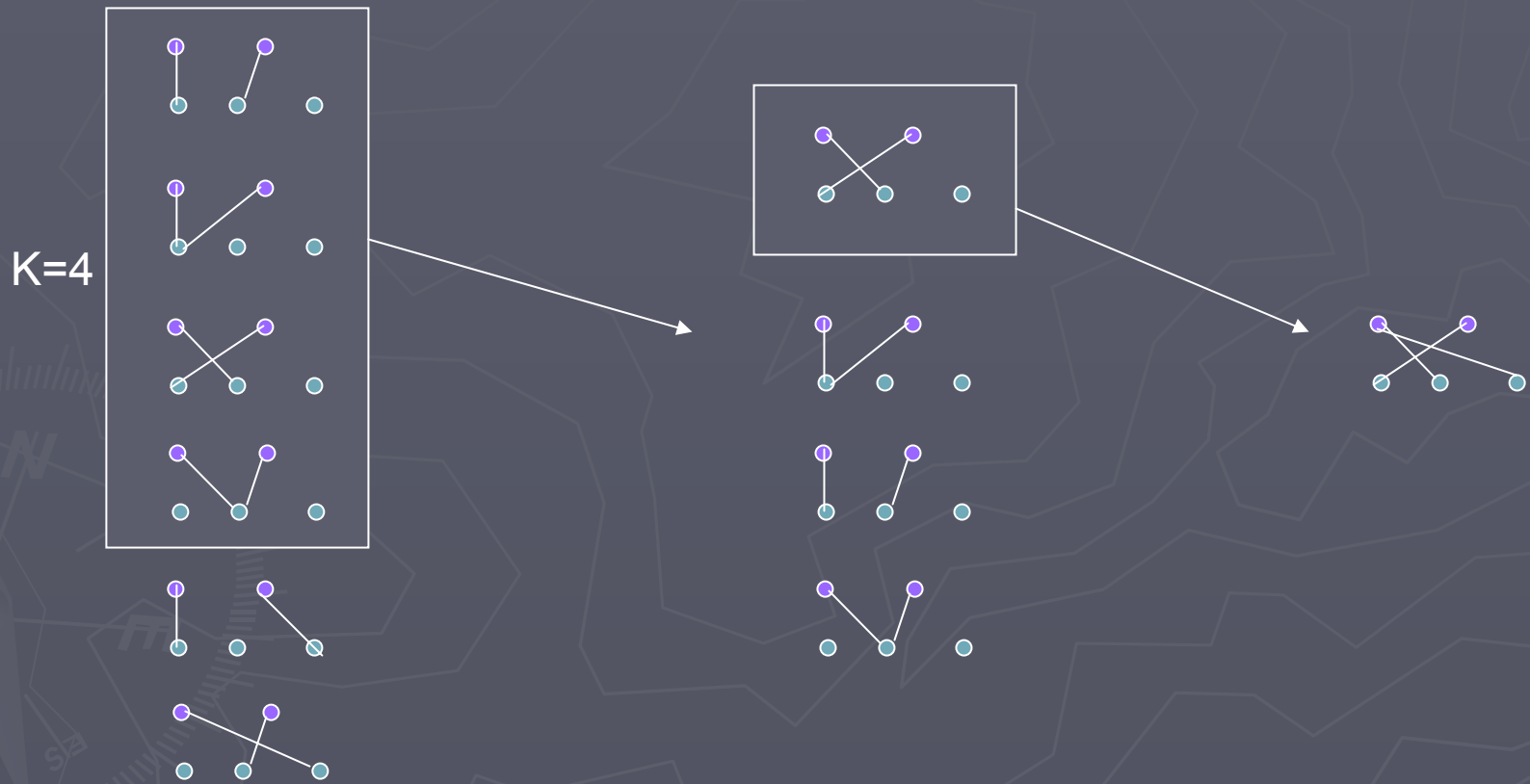
Discriminative Re-ranking Approach

- ▶ The best alignment $\hat{a} = \operatorname{argmax} \operatorname{score}(a | e, h)$
- ▶ Here, e is the english and h is the hindi sentence.
- ▶ $\operatorname{score}(a | e, h) = \operatorname{score}_{L_a}(a | e, h) + \operatorname{score}_S(a | e, h)$

Alignment search (Discriminative Re-ranking)

- ▶ Three main steps
 - Populate the Beam
 - ▶ Use local features to determine K-best alignments of source words with words in the target sentence.
 - Re-order the Beam
 - ▶ Re-order the above alignments using structural features.
 - Post-processing
 - ▶ Extend alignments to include other links that can be inferred using simple rules.

Alignment search (Discriminative Re-ranking)



Populate the Beam

- ▶ Obtain K-best candidate alignments using local scores .
- ▶ Local score is computed by looking at the features of the individual alignment links independently.
- ▶ $\text{score}_L(e_j, h_k) = W \cdot f_L(e_j, h_k)$
- ▶ $\text{score}_{La}(a | e, h) = \sum \text{score}_L(e_j, h_k)$

Populate the Beam - 2

- ▶ Task: Populate the beam in the decreasing order of $\text{score}_{L_a}(a | e, h)$.
- ▶ Compute the local score of each source word with every target word (including NULL).
- ▶ Top-k alignment links of each source word are chosen.

Populate the Beam - 2

- ▶ Populating K-best alignments
 - Implemented using Priority Queues.
- ▶ Initial State of Priority Queue
 - One entry representing the best alignment (set of best alignment links).
- ▶ At every iteration
 - Pop the best entry from the PQ.
 - Add it's k successor entries back into the PQ.

Re-order the Beam

- ▶ Structural scores are now added to the local scores of the alignments in the beam in order to re-order the beam.
 - $\text{score}_s(a) = W \cdot f_s(a)$
- ▶ Overall score = $\text{score}_{La}(a) + \text{score}_s(a)$
- ▶ Structural features look at properties of the entire alignment structure instead of individual alignment links.

Post-processing

- ▶ Previous two steps produce alignments which contain one-to-one and many-to-one mappings.
- ▶ Goal is to include the best alignment structure from previous step to include other alignment links of one-to-many/many-to-many types.
- ▶ New alignment links are added while processing source words in the breadth first order of the **dependency structure**.

Post-processing

- ▶ Algorithm:
- ▶ Let w be next word considered. $pw = \text{parent}(w)$.
 - If w, pw linked to one or more common words.
Align w to all words already aligned with pw .
 - Else, Use simple target-specific rules to extend alignments of w .
- ▶ Recursively consider all the children of w

Post-processing

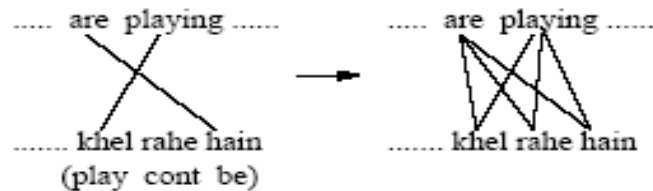


Figure 1: Inferring the many-to-many alignments of verb and auxiliaries

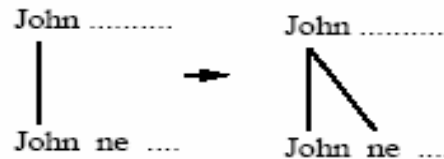


Figure 2: Inferring the one-to-many alignment to case-markers in Hindi

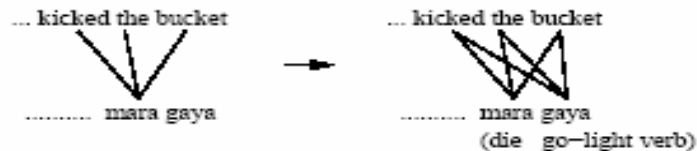


Figure 3: Inferring many-to-many alignment for source idioms

Outline

- ▶ Word Alignment - English-Hindi Language Pair
- ▶ Related approaches
- ▶ Discriminative Re-ranking approach
 - **Features**
 - Parameter optimization using MIRA
 - Results
- ▶ Future Work and Conclusion

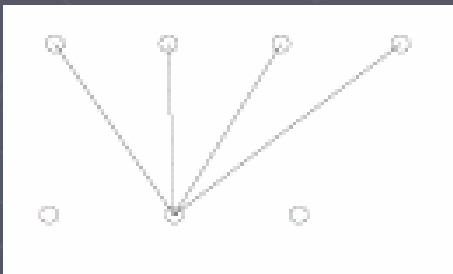
Features - Local

- ▶ DiceWords (Taskar et al., 2005)
- ▶ DiceRoots : Lemmatized forms of e_j and h_k .
- ▶ Dict : Whether there exists an entry from source word e_j to target word h_k .
- ▶ Null(POS) : Binary feature which is active when a source word with a particular part-of-speech tag is aligned with NULL.

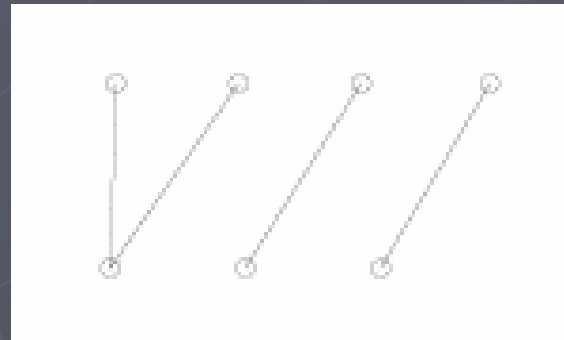
Structural Features

► Overlap

- This feature considers the instances in a sentence pair where a source word links to a target word which is a participant in more than one alignment link.



<



$$\text{Overlap}(\bar{a}) = \frac{\sum_{h_q \in T, \text{Fert}(h_q) > 1} \text{Fert}^2(h_q)}{\sum_{h \in T} \text{Fert}(h)}$$

Structural Features

► Null Percent

- This feature measures the percentage of words in target sentence with zero fertility.

$$\text{NullPercent} = \frac{|h_q |_{h_q \in T, \text{Fertility}(h_q) == 0}}{|h|_{h \in T}}$$

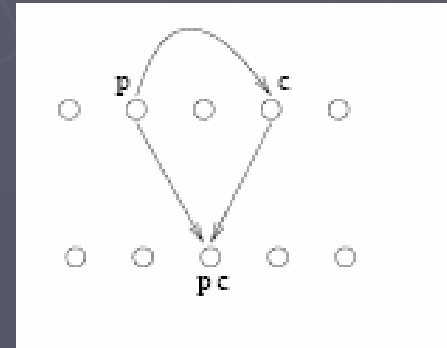
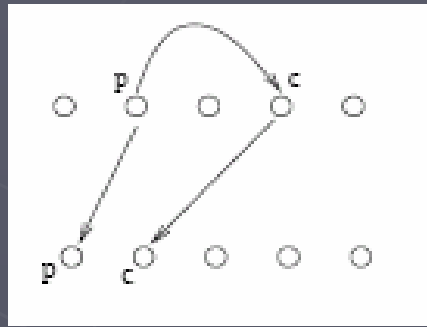
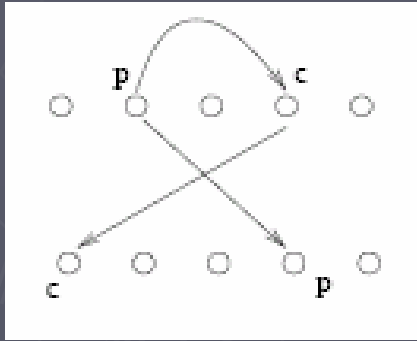
Structural Features

► Direction of Dependency Pair

- Captures first order interdependence between the alignments links connected to two sources connected by a dependency relation.
- One way to measure such interdependence is by noting the order of target sentence words the child and the parent of a source sentence dependency relation.
- Three possible orders (next slide).

Structural Features

► Direction of Dependency Pair



- The feature thus captures a simple divergence between the source and target dependency structures.

Outline

- ▶ Word Alignment - English-Hindi Language Pair
- ▶ Related approaches
- ▶ Discriminative Re-ranking approach
 - Features
 - **Parameter optimization using MIRA**
 - Results
- ▶ Future Work and Conclusion

Online large margin Training using MIRA

- ▶ For parameter optimization, we used online-large margin algorithm called MIRA (Crammer and Singer, 2005; McDonald et al., 2005).
- ▶ If $T = \{ (x_i, y_i) \}^m$ be gold data, where x_i is the i^{th} sentence pair, y_i is the corresponding gold alignment. The task is to learn the weight vector W such that,

Online large margin Training using MIRA

- ▶ For a sentence pair, the weight should be optimized in the following fashion.

Minimize $\| w_{i+1} - w_i \|$

Such that

$$w \cdot f(x_i, y_i) - w \cdot f(x_i, y'_i) \geq \text{loss}(y_i, y'_i)$$

For all, $(x_i, y_i) \in T$, $y'_i \in K\text{-best Predictions}(x_i)$

- ▶ Online training algorithm.

Outline

- ▶ Word Alignment - English-Hindi Language Pair
- ▶ Related approaches
- ▶ Discriminative Re-ranking approach
 - Features
 - Parameter optimization using MIRA
 - **Results**
- ▶ Future Work and Conclusion

Data

- ▶ Unsupervised data: 50,000 sentence pairs
- ▶ Supervised data
 - Training : 4252 sentence pairs
 - Testing : 100 sentence pairs



GIZA++ results

<i>Mode</i>	<i>Prec.</i>	<i>Rec.</i>	<i>F-meas.</i>	<i>AER</i>
<i>Normal: Eng-Hin</i>	47.57	40.87	43.96	56.04
<i>Normal: Hin-Eng</i>	47.97	38.50	42.72	57.28
<i>Normal: Inter.</i>	88.71	27.52	42.01	57.99
<i>Lemma.: Eng-Hin</i>	53.60	44.58	48.67	51.33
<i>Lemma.: Hin-Eng</i>	53.83	42.68	47.61	52.39
<i>Lemma.: Inter.</i>	86.14	32.80	47.51	52.49

Table 3: Giza++ Results

Results using local features

<u>Features</u>	<u>Precision</u>	<u>Recall</u>	<u>F-measure</u>	<u>AER</u>
Dicewords + Diceroots	41.49	38.71	40.05	59.95
+ Null_POS	42.82	38.29	40.43	59.57
+ Dict	43.94	39.30	41.49	58.51
+ Word pairs	46.27	41.07	43.52	56.48

Results after adding Global features

<u>Features</u>	<u>Precision</u>	<u>Recall</u>	<u>F-measure</u>	<u>AER</u>
Local feats.	46.27	41.07	43.52	56.48
Local feats. + Overlap	48.17	42.76	45.30	54.70
Local feats + Direct_Deppair	47.93	42.55	45.08	54.92
Local feats + All struct. feats	48.81	43.31	45.90	54.10

Adding structural features to Giza transition probabilities

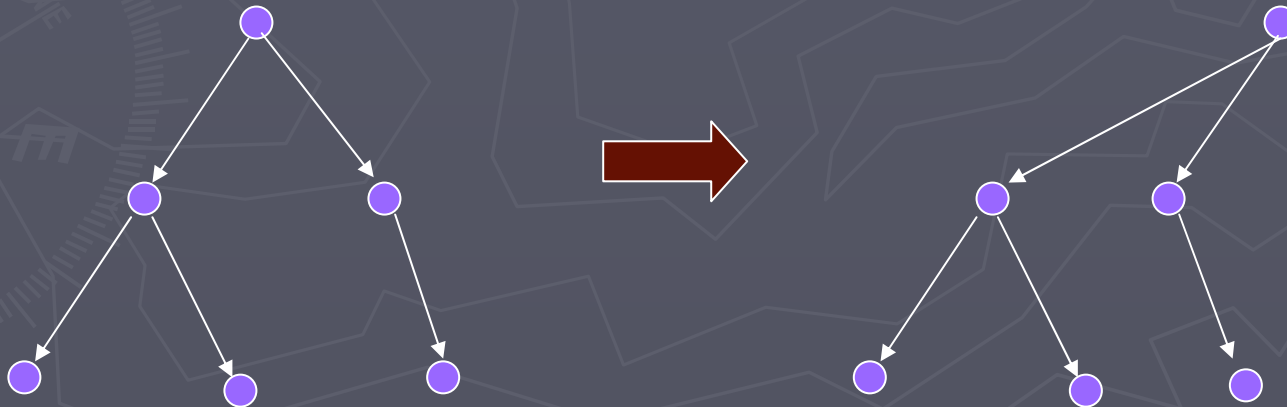
<u>Features</u>	<u>Precision</u>	<u>Recall</u>	<u>F-measure</u>	<u>AER</u>
IBM Model-4 Pars. + Local feats.	48.85	43.98	46.29	52.71
Local feats. + All struct. feats	48.95	50.06	49.50	50.50

Outline

- ▶ Word Alignment - English-Hindi Language Pair
- ▶ Related approaches
- ▶ Discriminative Re-ranking approach
 - Features
 - Parameter optimization using MIRA
 - Results
- ▶ Future Work and Conclusion

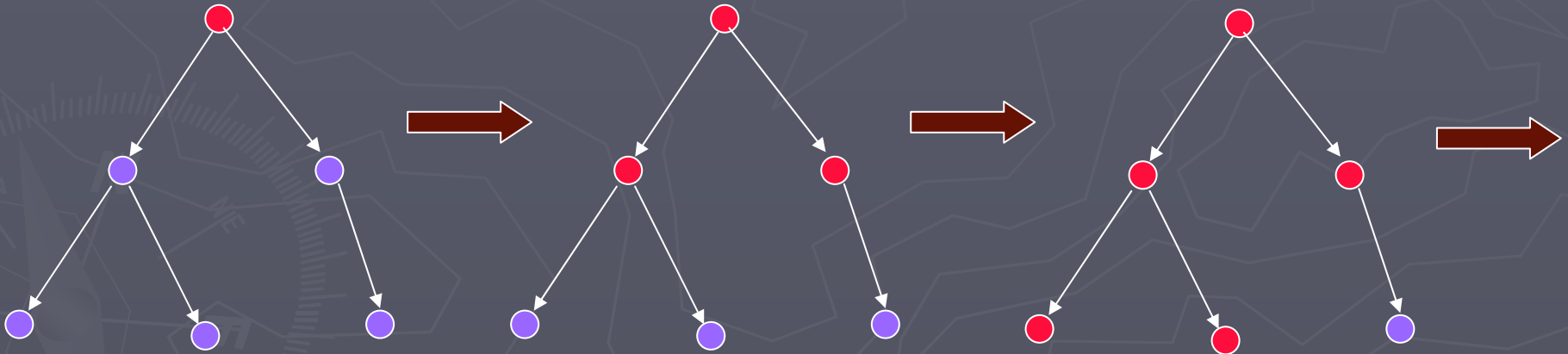
Future work

- ▶ Experiment with more sophisticated structural features.
- ▶ Design an transducer (dependency based) which uses parameter weights learnt by our approach and the LM.



Future work

- ▶ Merge the two alignment search steps to make better use of structural features.



The background of the slide is a dark grey color with a faint, light grey topographic map overlay. The map features various contour lines and a compass rose in the lower-left quadrant. The compass rose includes a needle pointing towards the top-left and is labeled with 'N' for North, 'E' for East, and 'S' for South. The text 'THANK YOU' is centered on the slide in a white, sans-serif font.

THANK YOU

Questions and Suggestions ?