

Translating Short Segments with NMT: A Case Study in English-to-Hindi



Shantipriya Parida and Ondřej Bojar
Charles University, Faculty of Mathematics and Physics
Institute of Formal and Applied Linguistics
Prague, Czech Republic
Email: {parida,bojar}@ufal.mff.cuni.cz



Introduction

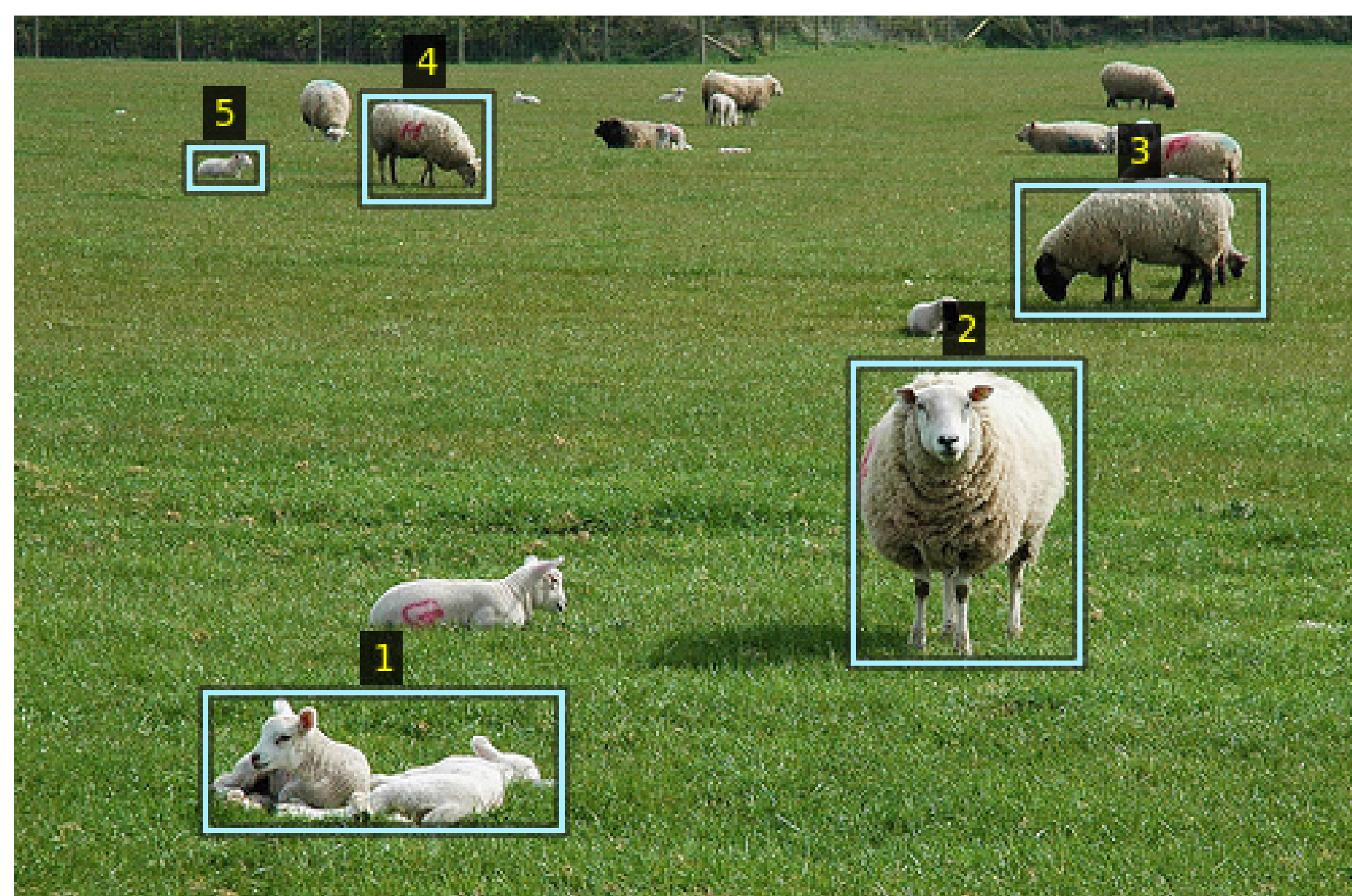
- **Visual Genome:** Dataset of images, captions and relations potentially useful for many text and image processing applications.
- 108k images with 5.4M short captions in English.

Motivation for Hindi Visual Genome

- The Hindi version of Visual Genome would allow researchers to study multi-modal NLP for the world's fourth most spoken language.
- *Parallel* to the English original, this resource would serve in multi-modal MT research.

- In this work: Set up a solid baseline MT.
- Next step: Find ambiguous segments where image or surrounding captions could help.

Context Disambiguates

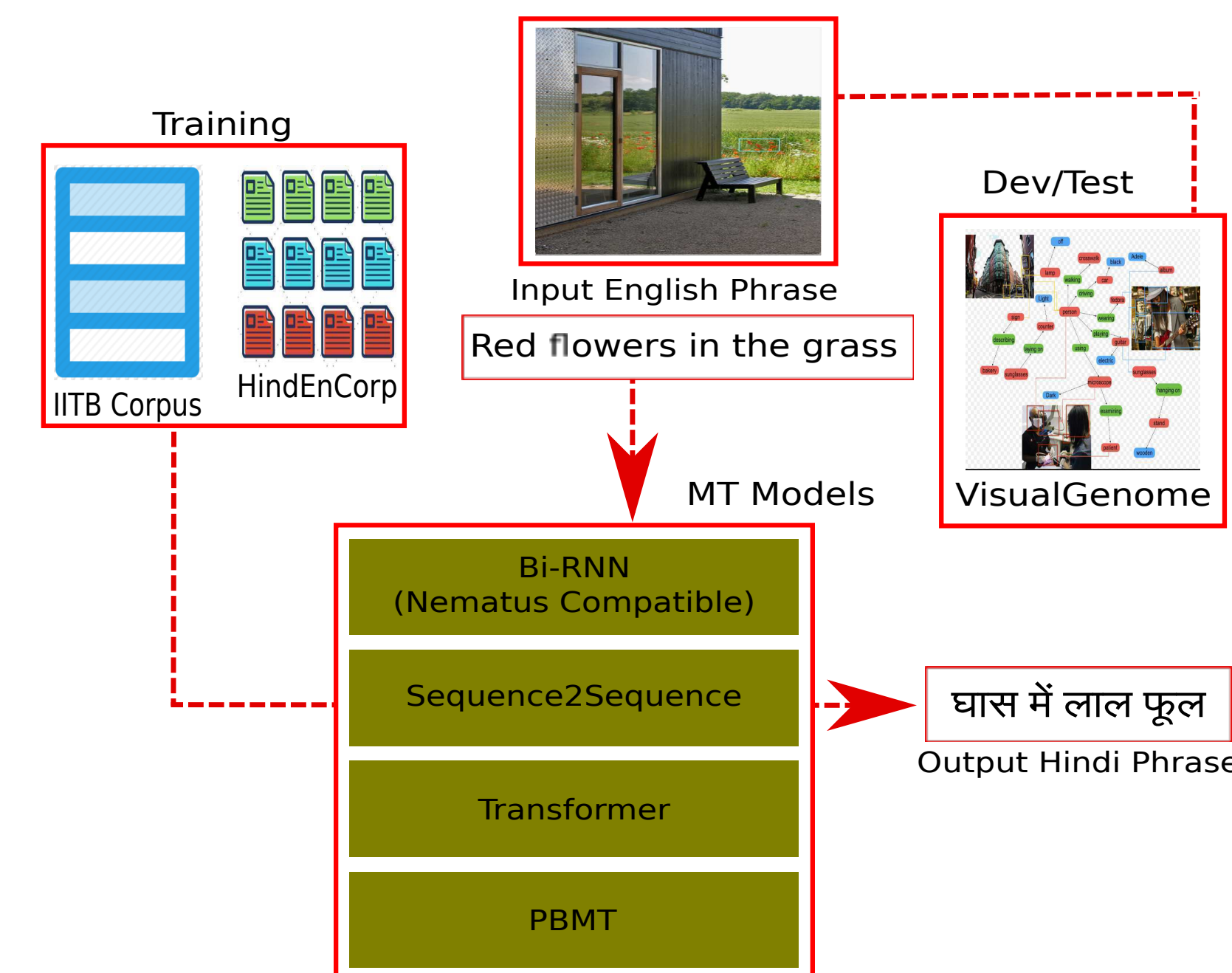


Caption 1: Two lambs lying in the sun.
Hindi MT: दो भेड़ के बच्चे सूरज में झूठ बोल रहे हैं
Gloss: Two baby sheep are **telling lies** in the sun.

Selected surrounding captions:

1. Sheep standing in the grass
2. Sheep with black face and legs
3. Sheep eating grass
4. Lamb sitting in grass.
5. Lamb sitting in grass.

Experiments



Training and Evaluation Data:

Dataset	#Sentences	#Tokens	En	Hi
Train (HindEnCorp)	273.9k	3.8M	5.6M	
Train (IITB)	1492.8k	20.8M	31.4M	
Dev (Visual Genome)	898	4519	6219	
Test (Visual Genome)	1000	4909	6918	

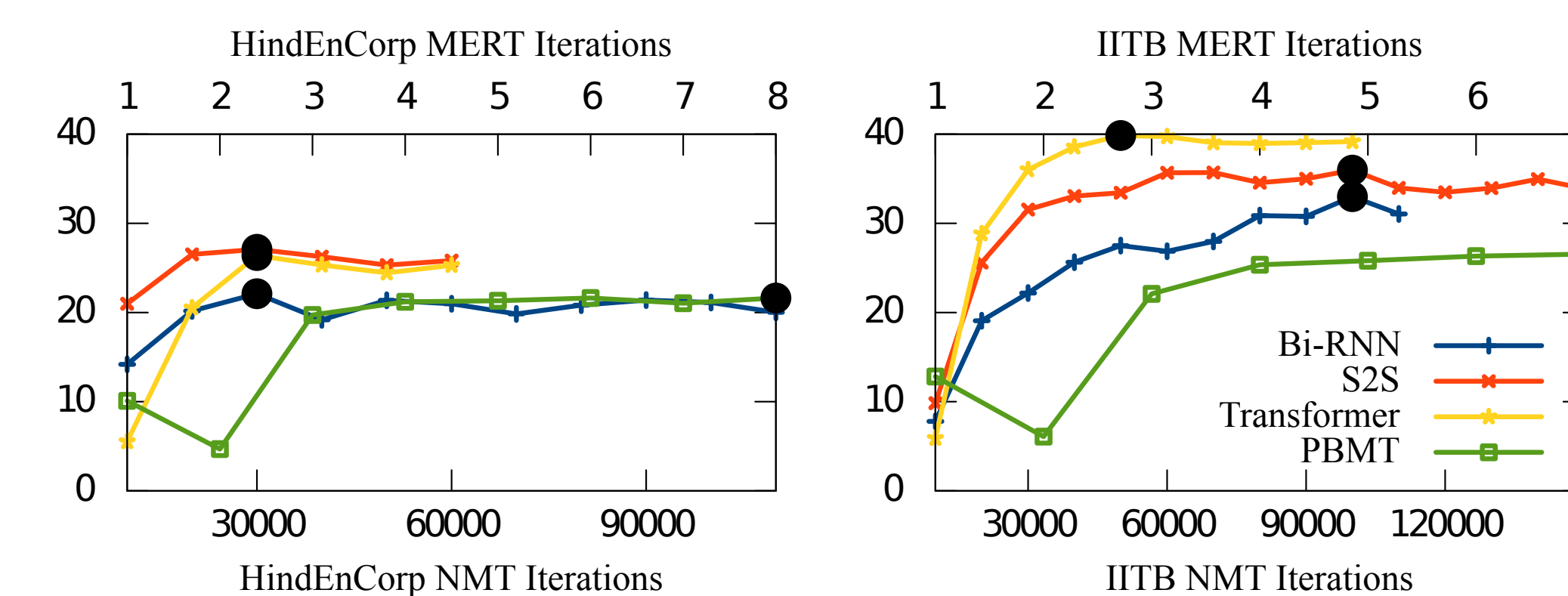
- **NMT Toolkit:** Marian (C++ implementation of several models).
- **MT Models** tested:
 - Marian's nematus Model (Bi-RNN), used shallow.
 - Marian's Sequence-to-Sequence (s2s) Model, used deep.
 - Marian's transformer Model.
- **Common Settings:** Tokenized with Moses tokenizer, joint BPE trained on HindEnCorp, 30k merge operations. Trained on four GeForce GTX 1080 Ti GPUs for 14 hours (best score).
- **Baseline:** Moses Phrase-Based MT with 5-gram language model.

Transformer is very sensitive to hyperparameters (requires more experimenting).

Parameter	Bi-RNN	S2S	Transformer	Parameter	Bi-RNN	S2S	Transformer
beam-size	12	12	12	enc-depth	1	4	6
dec-cell	gru	lstm	-	enc-type	bidirectional	alternating	-
dec-cell-base-depth	2	4	-	exponential-smoothing	-	0.0001	-
dec-cell-high-depth	1	2	-	heads	-	-	8
dec-depth	1	4	6	label-smoothing	-	-	0.1
decay-inv	-	-	16000	learning-rate	0.0001	0.0001	0.0003
dim-emb	512	512	512	max-length	50	50	100
dim-rnn	1024	1024	1024	normalize	-	-	0.6
dropout-rnn	0.2	0.2	-	optimizer	adam	adam	adam
dropout-src	0.1	0.1	-	transformer-dim-ffn	-	-	2048
dropout-trg	0.1	0.1	-	transformer-dropout	-	-	0.1
early-stopping	10	-	-	transformer-dropout-attention	-	-	0
enc-cell	gru	lstm	-	transformer-postprocess	-	-	dhc
enc-cell-depth	1	2	-	warm-up	-	-	16000

Results

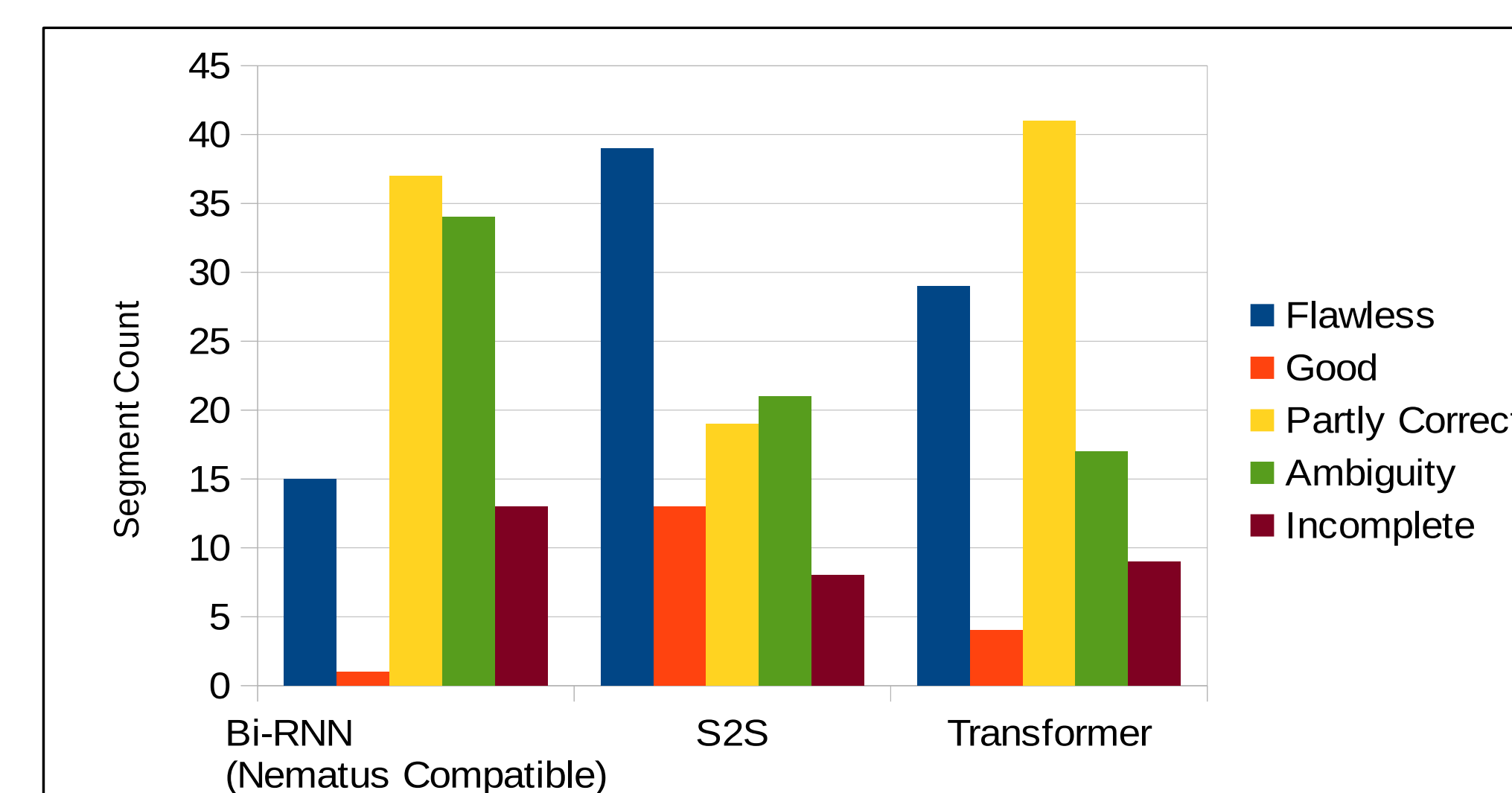
Transformer fastest and best in BLEU.



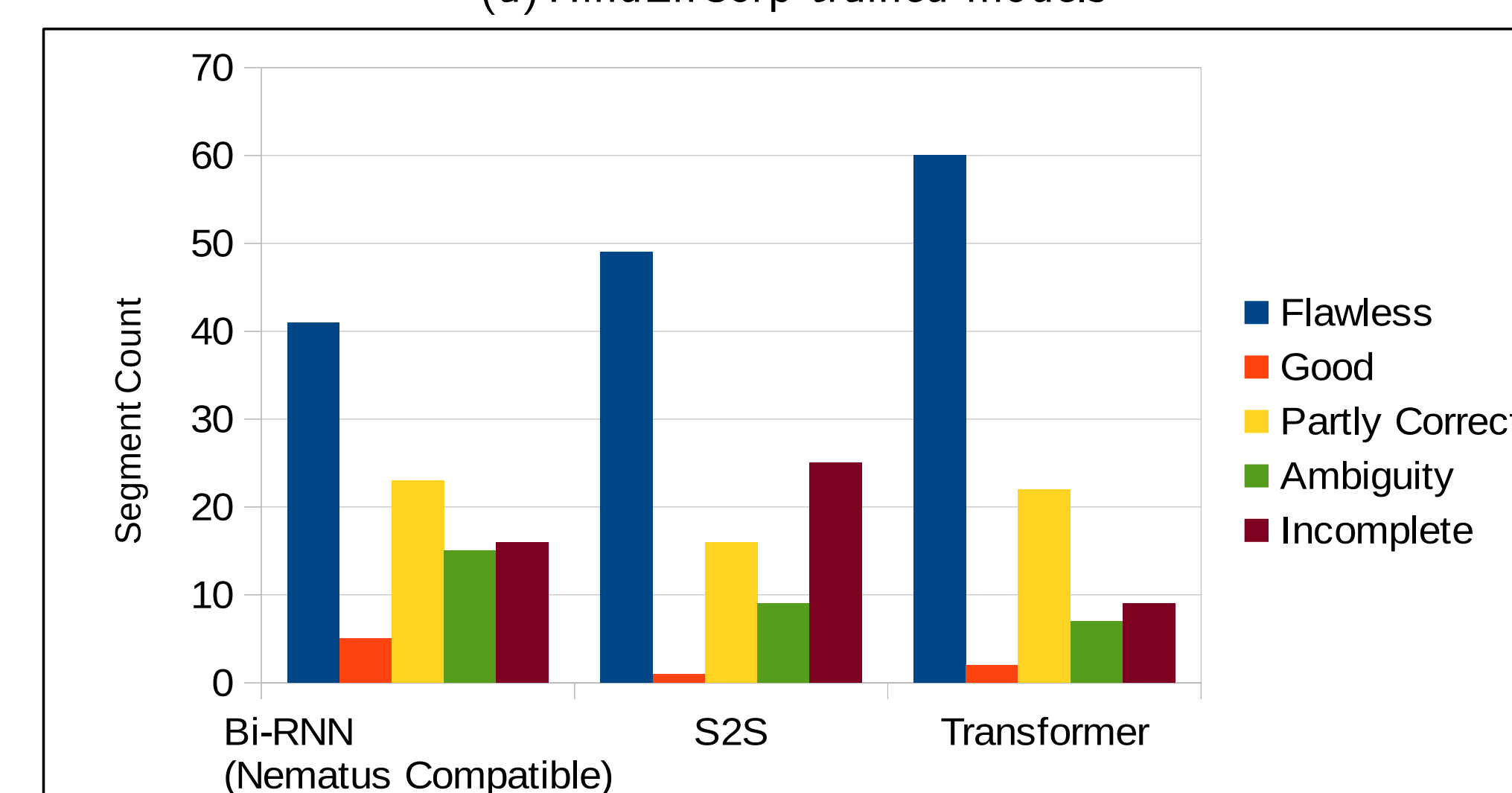
BLEU (dev set). Black dots indicate the iteration used for test set translation and evaluation.

Manual Evaluation

Deep S2S better for small data: more flawless outputs.



(a) HindEnCorp-trained models

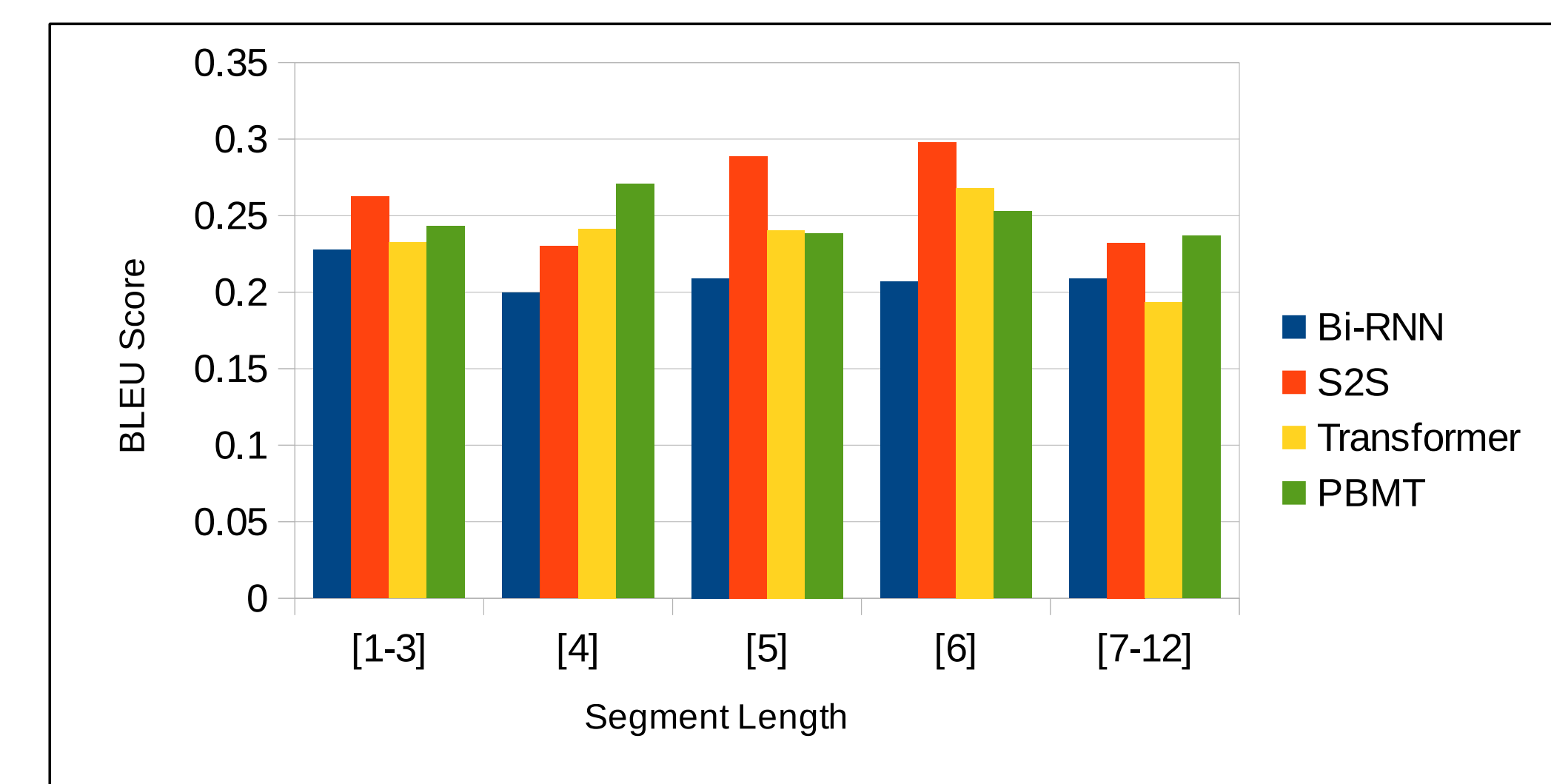


(b) IITB-trained models

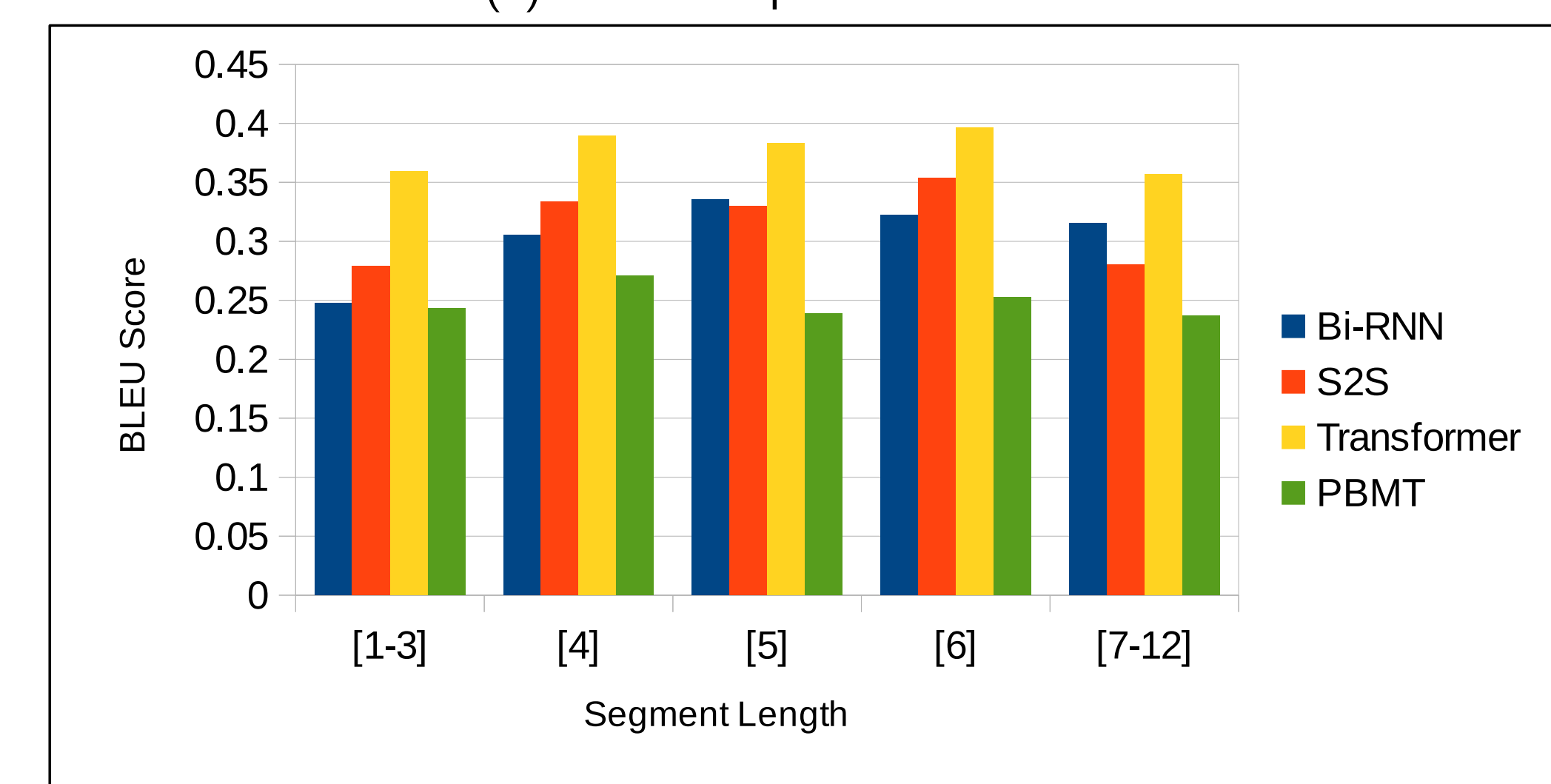
Further Analysis

No clear tendency in translation quality across on source lengths.

- PBMT a little better in small data setting (lengths of 1-3, 4, and 7+).
- Transformer wins for all lengths with large data.

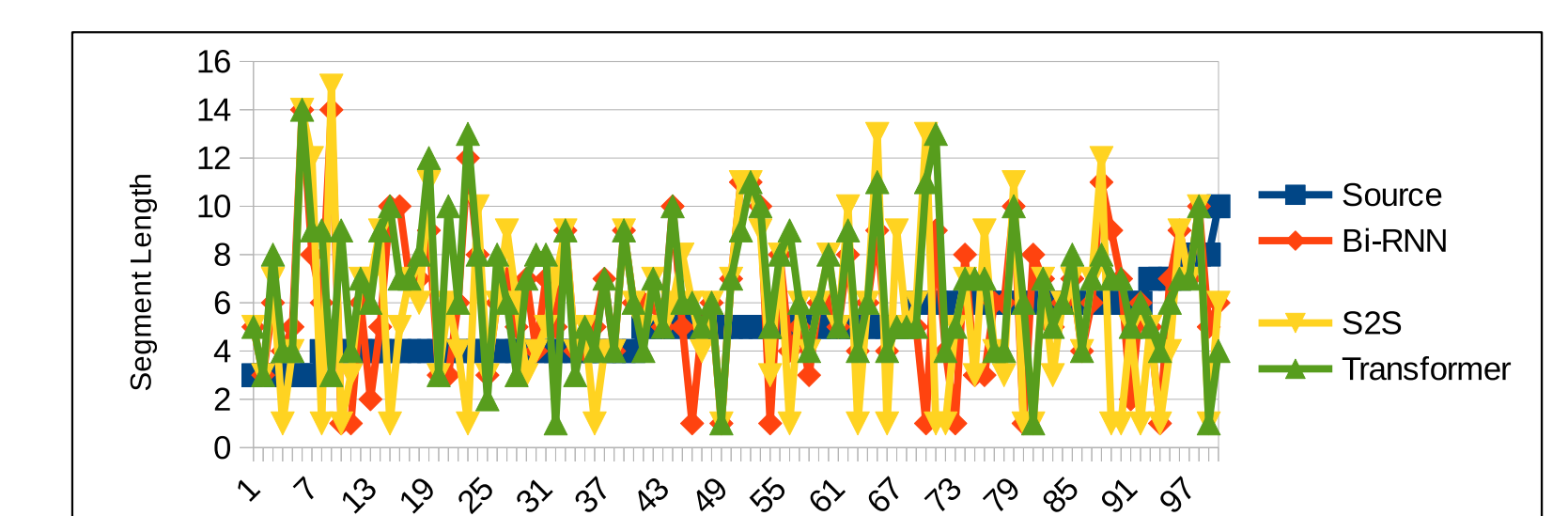


(c) HindEnCorp-trained models



(d) IITB-trained models

Target length varies across segments and NMT models.



Source and candidate translation lengths for individual segments (sorted by source length)

This work has been supported by the grants 18-24210S of the Czech Science Foundation, SVV 260 453 and "Progress" Q18+Q48 of Charles University, and using language resources distributed by the LINDAT/CLARIN project of the Ministry of Education, Youth and Sports of the Czech Republic (projects LM2015071 and OP VVV VI CZ.02.1.01/0.0/0.0/16 013/0001781).