Addressing Word-order Divergence in Multilingual Neural Machine Translation for Extremely Low Resource Languages

Rudra Murthy [1] Anoop Kunchukuttan [2] Pushpak Bhattacharyya [1]

[1]Indian Institute Of Technology Bombay, Mumbai, India [2]Microsoft AI & Research, Hyderabad, India

Overview

Goal:

- Does word-order divergence affect Multilingual Neural Machine Translation?
- Improve MT performance between source language (**A**) and target language (**B**) (**A** → **B**)

Assumptions:

- Very less or no parallel corpus exists between languages **A** and **B**
- However, very large parallel corpus exists between languages **C** and **B**
- Assisting source language (**C**) has different word-order compared to source language (**A**)

Contribution:

- Show word-order divergence affects Multilingual Neural Machine Translation
- Pre-ordering assisting source language (**C**) sentences to match the word-order of source language (**A**) sentences leads to better Multilingual NMT performance
- Beneficial for extremely low-resource languages with very small or no parallel data

Motivation

- Language divergence could however negate the benefits from multilingual learning leading to drop in performance
- Specifically, the word-order divergence between *assisting-source* and *source* languages

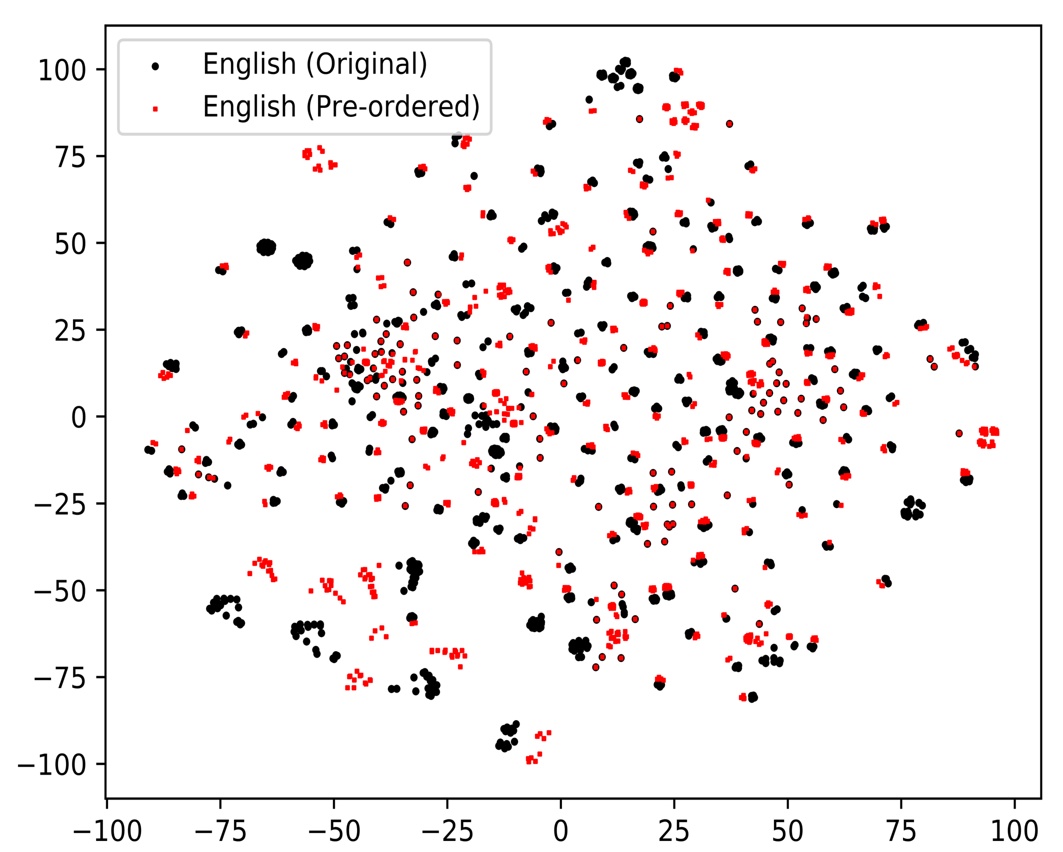
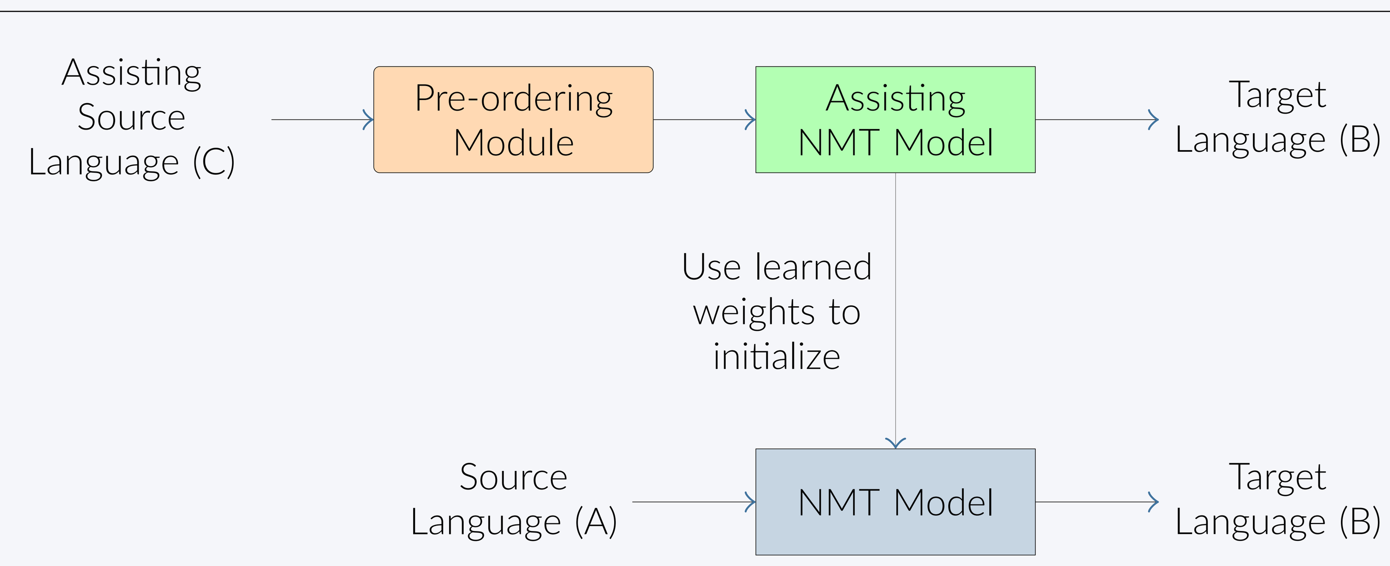
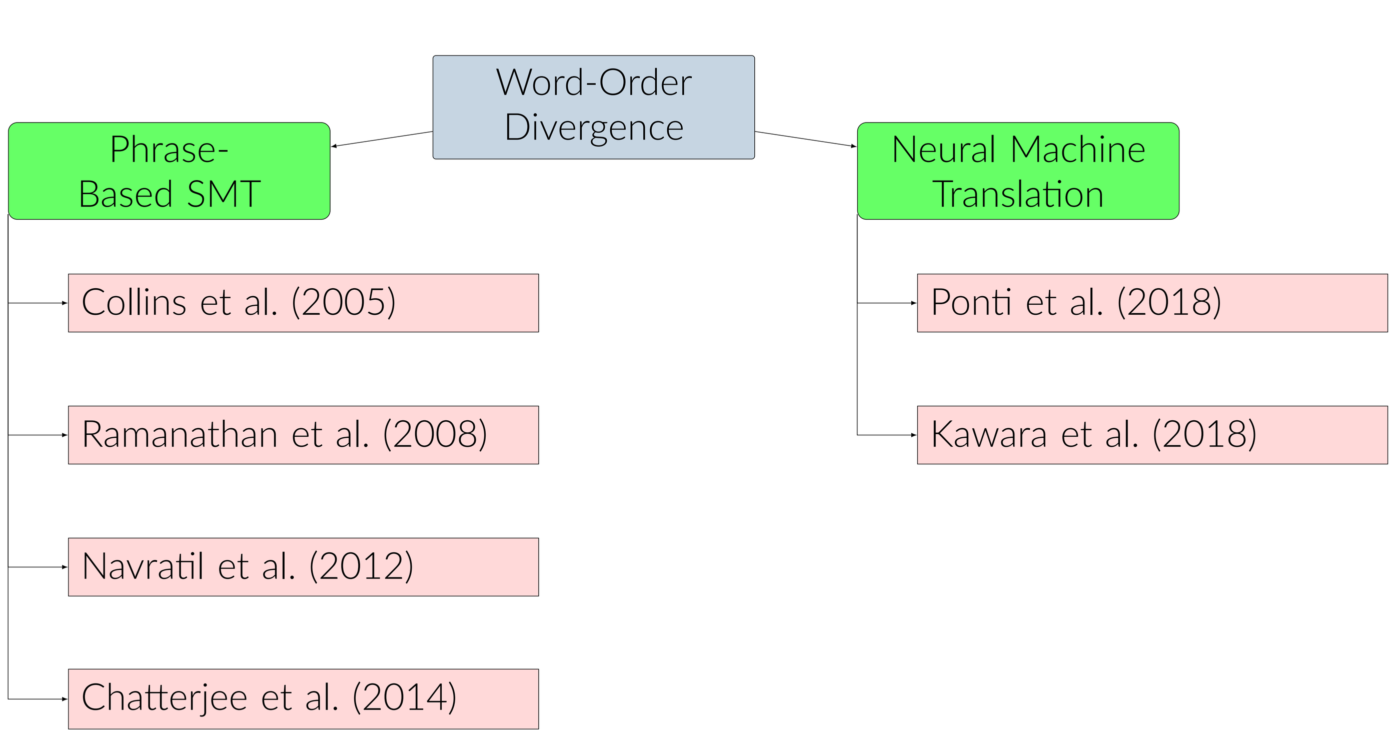


Figure 1: Encoder Representations for English sentences with and without Pre-Ordering

Approach



Related Work



We address word-order divergence between source languages as opposed to previous works which addressed divergence between source and target languages.

Experimental Setup

Task	Indian Languages → Hindi
Source Languages	1. Bengali 2. Gujarati 3. Marathi 4. Malayalam 5. Tamil
Assisting Source Language	English (SVO)
	Corpus Size
English → Hindi	1.46 M sentences (Kunchukuttan et al. (2018))
Indian Languages → Hindi	44.7 K sentences (Jha (2010))
Test Data Size	2K sentences (Jha (2010))
Network	Experiment with both Bi-LSTMs and Transformer networks
Pre-Ordering Rules	Generic Pre-Ordering Rules (G) (Ramanathan et al. (2008)) Hindi-Tuned Pre-Ordering Rules (HT) (Patel et al. (2013))
Word Embeddings	English Fasttext embeddings Word Translate source language data to English using Google Translate

Results

Pre-ordering performs better than no pre-ordered system using Bi-LSTM encoder decoders.

Language	BLEU			LeBLEU (%)		
	No Pre-Order	Pre-Ordered		No Pre-Order	Pre-Ordered	
		HT	G		HT	G
Bengali	6.72	8.83	**9.19**	37.10	41.50	**42.01**
Gujarati	9.81	**14.34**	13.90	43.21	47.36	**47.60**
Marathi	8.77	10.18	**10.30**	40.21	41.49	**42.22**
Malayalam	5.73	6.49	**6.95**	33.27	33.69	**35.09**
Tamil	4.86	**6.04**	6.00	29.38	30.77	**31.33**

Table 1: Transfer learning results for *X-Hindi* pair using Bi-LSTM model, trained on *English-Hindi* corpus and sentences from *X* word translated to English. HT: Hindi-tuned Pre-ordering Rules, G: Generic Pre-ordering Rules

Pre-ordering performs better than no pre-ordered system even when Transformer networks are used.

Language	Positional Encoding			No Positional Encoding		
	No Pre-Order	Pre-Ordered		No Pre-Order	Pre-Ordered	
		HT	G		HT	G
Bengali	6.03	**8.61**	8.16	4.07	**4.52**	4.05
Gujarati	8.43	**12.20**	11.01	4.94	**5.71**	5.08
Marathi	6.96	**9.16**	8.68	4.40	**5.07**	5.03
Malayalam	4.37	**5.69**	5.08	3.56	**4.08**	3.63
Tamil	3.89	**5.08**	4.64	2.71	**3.14**	2.82

Table 2: Transfer learning results (BLEU Score) for *X-Hindi* pair using Transformer networks, trained on *English-Hindi* corpus and sentences from *X* word translated to English. HT: Hindi-tuned Pre-ordering Rules, G: Generic Pre-ordering Rules

When *Source-Target* parallel corpus is available, the fine-tuned (no pre-ordered) model performs almost as good as the pre-ordered model

Corpus Size	No Transfer Learning	No Pre-Order	Pre-Ordered	
			HT	G
Bengali				
-	-	6.72	8.83	**9.19**
500	0.0	11.40	**11.49**	11.00
1000	0.0	13.71	**13.84**	13.62
2000	0.0	16.41	**16.79**	16.01
3000	0.0	17.44	**18.42**†	17.82
4000	0.0	18.86	**19.17**	18.66
5000	0.07	19.58	**20.15**†	19.82
10000	1.87	22.50	**22.92**	22.53

Corpus Size	No Transfer Learning	No Pre-Order	Pre-Ordered	
			HT	G
Gujarati				
-	-	9.81	**14.34**	13.90
500	0.0	17.27	17.11	**17.75**
1000	0.0	21.68	**22.12**	21.45
2000	0.0	25.34	**25.73**	25.63
3000	0.29	27.48	27.77	**27.83**
4000	0.82	29.20	29.49	**29.51**
5000	0.0	29.87	**31.09**†	30.58†
10000	1.52	33.97	**34.25**	34.08

Table 3: Transfer learning results (BLEU Score) using Bi-LSTM model for *Indian Language-Hindi* pair, fine-tuned with varying number of *Indian Language-Hindi* parallel sentences. †Indicates statistical significance between *Pre-ordered* and *No Pre-ordered* results.

English	the treatment of migraine is done in two ways								
Gujarati (Original)	માઈગ્રેનની	સારવાર	બે	રીતે	કરી	શકાય	છે.		
Gujarati (Word Translate)	migraine	treatment	two	the way	doing be	done	is there .		
Hindi (Reference)	माइग्रेन	का	ट्रीटमेंट	दो	तरह	से	किया	जाता	है ।
(Word Translate)	migraine	of	treatment	two	kind	from	did	go	is .
No Pre-Order	<unk> <unk>	उपचार upachAra treatment	दो do two	प्रकार prakAra kind	से se from	किया kiyA did	जाता jAtA go	है hai is	। . .
Pre-ordered (HT)	माइग्रेन mAigrena migraine	का kA of	उपचार upachAra treatment	दो do two	तरह prakAra kind	से se from	किया kiyA did	जाता jAtA go	है। hai. is .

Table 4: Sample output generated by our Gujarati-Hindi NMT model. Text in red: phrases dropped by the no pre-ordered model.

Conclusion

- Pre-ordering the assisting language to match the word order of the source language significantly improves translation quality in an extremely low-resource setting.
- Alternatively, fine-tuning on a small source-target parallel corpus is sufficient to overcome word order divergence.

Future Work

- Validate the hypothesis on a more diverse set of languages and other word-order divergence Tarscenarios.
- Alternative methods to address word-order divergence which do not require expensive parsing.
- Apply to other multilingual NLP problems.

References

Chatterjee, R., Kunchukuttan, A., and Bhattacharyya, P. (2014). Supertag based pre-ordering in machine translation. In *Proceedings of the 11th International Conference on Natural Language Processing, ICON 2014*.

Collins, M., Koehn, P., and Kucerova, I. (2005). Clause restructuring for statistical machine translation. In *Proceedings of the 43rd Annual Meeting of the Association for Computational Linguistics (ACL'05)*. Association for Computational Linguistics.

Jha, G. N. (2010). The tdil program and the indian langauge corpora intitiative (ilci). In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*. European Languages Resources Association (ELRA).

Kawara, Y., Chu, C., and Arase, Y. (2018). Recursive neural network based preordering for english-to-japanese machine translation. In *Proceedings of ACL 2018, Student Research Workshop*. Association for Computational Linguistics.

Kunchukuttan, A., Mehta, P., and Bhattacharyya, P. (2018). The iit bombay english-hindi parallel corpus. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC-2018)*. European Languages Resources Association (ELRA).

Navratil, J., Visweswariah, K., and Ramanathan, A. (2012). A comparison of syntactic reordering methods for english-german machine translation. In *Proceedings of COLING 2012*. The COLING 2012 Organizing Committee.

Patel, R. N., Gupta, R., Pimpale, P. B., and M, S. (2013). Reordering rules for english-hindi smt. In *Proceedings of the Second Workshop on Hybrid Approaches to Translation*. Association for Computational Linguistics.

Ponti, E. M., Reichart, R., Korhonen, A., and Vulić, I. (2018). Isomorphic transfer of syntactic structures in cross-lingual nlp. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Association for Computational Linguistics.

Ramanathan, A., Hegde, J., Shah, R. M., Bhattacharyya, P., and M, S. (2008). Simple syntactic and morphological processing can help english-hindi statistical machine translation. In *Proceedings of the Third International Joint Conference on Natural Language Processing: Volume-I*.

Contact: {rudra,pb}@cse.iitb.ac.in, ankunchu@microsoft.com