Article

Dual-Source Transformer Model for Neural Machine Translation with Linguistic Knowledge

Yirong Pan [1,2,3], **Xiao Li** [1,2,3,*], **Yating Yang** [1,2,3,*] and **Rui Dong** [1,2,3]

[1] Xinjiang Technical Institute of Physics & Chemistry, Chinese Academy of Sciences
[2] University of Chinese Academy of Sciences
[3] Xinjiang Laboratory of Minority Speech and Language Information Processing
[*] Correspondence: xiaoli@ms.xjb.ac.cn, yangyt@ms.xjb.ac.cn

Abstract: Incorporating source-side linguistic knowledge into the neural machine translation (NMT) model has recently achieved impressive performance on machine translation tasks. One popular method is to generalize the word embedding layer of the encoder to encode each word and its linguistic features. The other method is to change the architecture of the encoder to encode syntactic information. However, the former cannot explicitly balance the contribution from the word and its linguistic features. The latter cannot flexibly utilize various types of linguistic information. Focusing on the above issues, this paper proposes a novel NMT approach that models the words in parallel to the linguistic knowledge by using two separate encoders. Compared with the single encoder based NMT model, the proposed approach additionally employs the knowledge-based encoder to specially encode linguistic features. Moreover, it shares parameters across encoders to enhance the model representation ability of the source-side language. Extensive experiments show that the approach achieves significant improvements of up to 2.4 and 1.1 BLEU points on Turkish→English and English→Turkish machine translation tasks, respectively, which indicates that it is capable of better utilizing the external linguistic knowledge and effective improving the machine translation quality.

Keywords: linguistic knowledge; neural machine translation model; machine translation tasks; knowledge-based encoder; model representation ability

1. Introduction

Neural machine translation (NMT) models based on encoder-decoder architecture are widely used for high-resource language pairs [1-5]. The NMT model employs the encoder to map the source sentence to a dense representation vector. Then it feeds the resulting vector to the decoder to produce the desired target sentence. In recent years, by exploiting the advanced neural network mechanisms, such as gating [2] and attention [3], the NMT model surpasses the previously dominant statistical machine translation (SMT) model [6] and achieves the state-of-the-art performance on many well-known machine translation tasks.

Instead of explicitly modeling the linguistic features in the SMT model, the NMT model directly learns the translation knowledge from the bilingual parallel sentences. Shi *et al.* indicated that the encoder captures both the local and global syntactic information of the source language, and different types of source syntax store in different layers with varying concentration degrees [7]. However, they also pointed out that the NMT model cannot capture the in-depth syntactic or semantic information of the source-side language. Therefore, it is difficult to perform machine translation tasks on the low-resource language pairs, especially for the morphologically-rich languages with a considerable vocabulary and complex morphology.

For Turkish→English machine translation task, the main problem is that the bilingual parallel sentences are far from sufficient, which cannot provide enough linguistic information for the NMT model to effectively learn the relationship between these two distinct languages. Moreover, the source-side Turkish is a morphologically-rich language with derivational morphology [8]. It has a

large vocabulary even in the low-resource training corpus, which makes great difficulty in translation tasks and leads to many inaccurate translation results [9-10]. For English → Turkish machine translation task, many English words have the same surface form with different word types and meanings, which corresponding to different translations. We consider that the external linguistic knowledge is benefit for the NMT model to enhance its representation ability on the source-side language. The first one is lemma, which is widely used for the information retrieval tasks. Since the vocabulary of morphologically-rich language is large, lemmatization can make better generalization by allowing the inflected and morphological variants of the same word to share their representations. The second one is part-of-speech (POS) tag. It can provide the syntactic role for each word in the context, which is helpful in extracting information and reducing data ambiguity. The third one is morphological tag. Since different word types have diverse sets of morphological features, morphology analysis can effectively reduce data sparseness.

Recently, incorporating source-side linguistic knowledge into the NMT model has achieved impressive performance on machine translation tasks. One popular method proposed by Sennrich and Haddow is to generalize the word embedding layer of the encoder to encode each word and its linguistic features without changing the other parts of the conventional NMT model [11]. However, since their method reduces the word embedding size to accommodate the additional input features, the model lacks an effective mechanism to balance the contribution from the word and its linguistic features. Instead, we propose a NMT approach that models the word and its linguistic knowledge, respectively. The other method proposed by Currey and Heafield is to change the architecture of the encoder to encode syntactic information [12]. They incorporates the source-side syntax into the NMT model by employing two separate encoders for the word sequence and its linearized parse. However, since their method combines the vector representations of word and syntax on the sentence-level, it lacks the flexibility to utilize other types of linguistic information. Instead, we employ the additional encoder to specially incorporate various linguistic features on the word-level.

Focusing on the above limitations, this paper proposes a novel NMT approach that models the words in parallel to the linguistic knowledge by using two separate encoders. Compared with the single encoder based NMT model, the proposed approach follows the multi-source framework [13] that additionally employs the knowledge-based encoder to specially encode linguistic knowledge. This is achieved by generalizing its word embedding layer to allow for an arbitrary number and type of linguistic features. Moreover, the proposed approach shares all parameters across encoders to enhance the model representation ability of the source-side language. Experimental results show that the approach achieves significant improvements in both Turkish→English and English→Turkish machine translation tasks.

2. Related Work

Many researchers show great interest in explicitly utilizing source-side syntactic and linguistic information as prior knowledge to improve their models. Eriguchi *et al.* modified the NMT model by building a tree-based encoder following the source-side parsed tree [14]. The model has an attention mechanism that allows the encoder to align both the input words and the input phrases with the output words. Yang *et al.* extended this work by encoding each node of the syntactic tree with both the local and global context information, and presented a weighted variant of the attention model to adjust the proportion of the conditional information [15]. Li *et al.* linearized the source-side syntactic structure into a label sequence and combined it with the word sequence, which makes the model automatically learning useful information [16]. Aqlan *et al.* integrated linguistic features on top of the word surface form in the translation model of the SMT, and iteratively trained the model to find the most optimized parameters [17]. Li *et al.* introduced a knowledge-aware approach that jointly models the word and the linguistic features, and added an attention gate to control the contributions from each encoder [18].

Multi-source NMT model is firstly proposed by Zoph and Knight for multilingual translation [13]. It is a *many-to-one* setting in the multi-task learning (MTL) approach [19]. The model consists of multiple encoders with one encoder per source language and combines the resulting sentence representations before feeding them into the decoder. Based on the multi-source NMT model, Currey and Heafield

incorporated the source-side syntax into the NMT model by employing two separate encoders, one encoder for the word sequence and the other encoder for the linearized parse [12]. Junczys-Dowmunt and Grundkiewicz employed a Transformer model with two encoders that ties word embeddings and shares parameters across encoders for automatic post-editing task [20].

Moreover, many semantic paring tasks also benefit from linguistic knowledge and multi-source framework. Liu *et al.* used the linguistic information of lemma under the conventional NMT model [21]. Duong *et al.* utilized multi-source NMT model with multiple encoders to represent the languages [22]. Noord *et al.* investigated to exploit linguistic knowledge in a multi-encoder setup for neural semantic parsing [23]. Different from previous models that simply combine the additional linguistic features, the proposed approach provides a reasonable and flexible way to incorporate the external linguistic knowledge into the NMT model. It generalizes the word embedding layer of the knowledge-based encoder to allow for an arbitrary number and type of linguistic features to enrich their representations.

3. Dual-Source Transformer Model with Linguistic Knowledge

3.1. Transformer Model

The Transformer model with self-attention mechanism is used in the proposed approach [5]. The model consists of an encoder and a decoder. The encoder maps the source sequence $x = (x_1, \ldots, x_m)$ to a continuous representation vector $z = (z_1, \ldots, z_m)$ by employing the multi-head attention component. And then the decoder produces the target sequence $y = (y_1, \ldots, y_n)$ based on all the previously generated symbols, the representation vector z, and a attention model.

3.2. Knowledge-based Encoder

Inspired by the NMT model proposed by Sennrich and Haddow [11], we generalize the input embedding layer of the knowledge-based encoder to incorporate external linguistic features into the NMT model. The architecture of the knowledge-based encoder is shown in Figure 1.
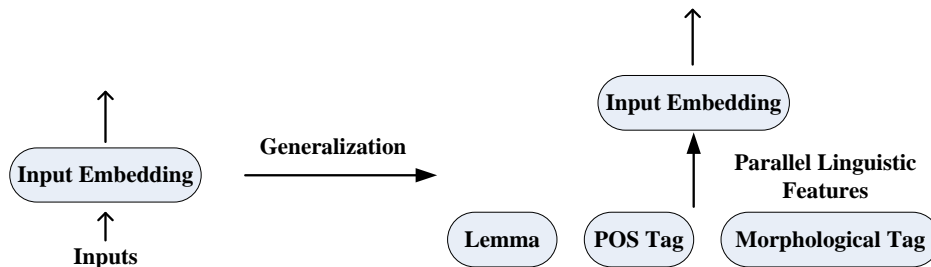

Figure 1. The architecture of the knowledge-based encoder.

Given $|F|$ linguistic feature sequences $k_1 = (k_{11}, \ldots, k_{1m})$, $k_2 = (k_{21}, \ldots, k_{2m})$, …, $k_{|F|} = (k_{|F|1}, \ldots, k_{|F|m})$ corresponding to the source sequence $x = (x_1, \ldots, x_m)$, the knowledge-based encoder maps k_1, k_2, …, $k_{|F|}$ to a representation vector $p = (p_1, \ldots, p_m)$, where p_i is computed by:

$$p_i = tanh \left(W \left(\bigcup_{t=1}^{|F|} E_t k_{ti} \right) + U p_{i-1} \right) \tag{1}$$

where $\cup$ is a vector concatenation operator, E_t is the feature embedding matrix, W and U are weight matrices. The length of the concatenated vector in the knowledge-based encoder equals to the source word embedding size.

3.3. Dual-Source Transformer Model

Inspired by the work of Zoph and Knight [13], the multi-source NMT model is employed to combine the vector representations of the word sequence and its corresponding linguistic features.

Since the proposed model consists of two separate encoders, it can be treated as a dual-source model. The architecture of the proposed dual-source Transformer model is shown in Figure 2.
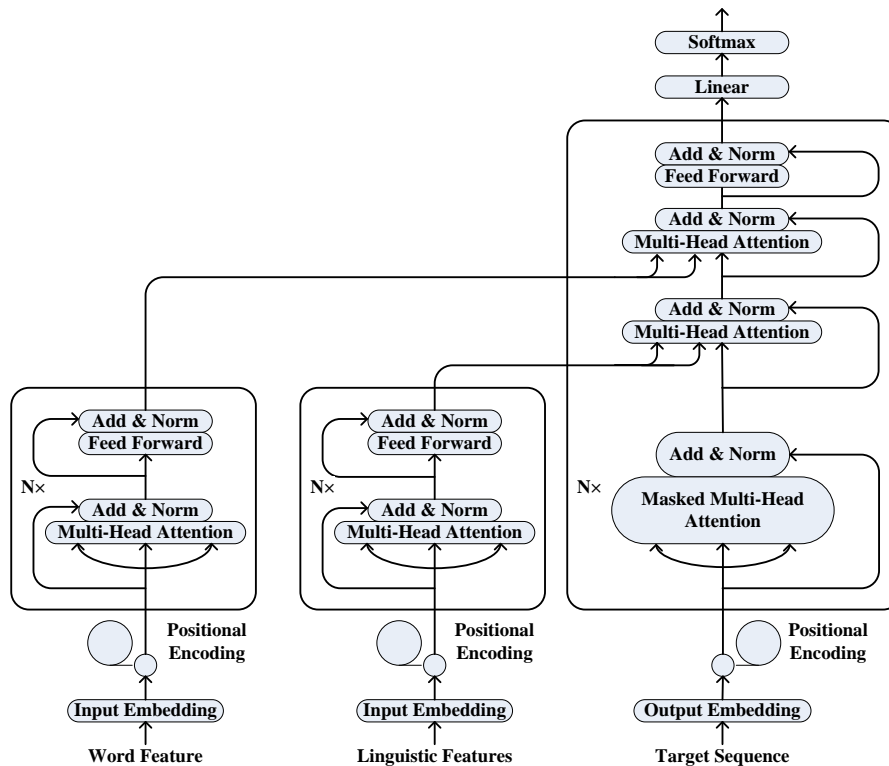

Figure 2. The architecture of the proposed dual-source Transformer model.

A combination method is used to concatenate the two hidden states from each encoder. Then, a linear transformation and a tanh nonlinear transformation are applied on the resulting vectors:

$$h = tanh\ (\boldsymbol{W_c}[h_1; h_2]) \tag{2}$$

where $\boldsymbol{W_c}$ is a weight matrice. The cell state is the sum of the two cell states from each encoder:

$$c = c_1 + c_2 \tag{3}$$

4. Linguistic Knowledge for Turkish and English

For Turkish→English machine translation task, the java toolkit Zemberek[1] with morphological disambiguation [24] is utilized to annotate the Turkish linguistic features of lemma, POS tag and morphological feature. Each word's morphological features is concatenated as its morphological tag. For English→Turkish machine translation task, the python toolkit NLTK[2] is utilized to annotate the English linguistic features of lemma and POS tag.

We use the BPE technique [25] to segment both the words and its lemmas into sub-word units, and add "@@" behind each non-final sub-word. We annotate the segmented lemma sequence with other linguistic features by coping the original word's feature value to all the sub-word units of its lemma. All the linguistic feature sequences have the same length. Examples of the word sequence and its corresponding linguistic feature sequences in the proposed dual-source model for Turkish and English are shown in Table 1 and Table 2, respectively.

Table 1. The sentence examples in the proposed dual-source model for Turkish.

Turkish Word Sequence	Ve bunlar sinek@@ kap@@ an an@@ em@@ onlar.						
Lemma Sequence	ve	bu	sinek@@	kapan	ane@@	mon	.

POS Tag Sequence	Conj	Pronoun	Noun	Noun	Noun	Noun	Punc
Morphological Tag Sequence	<null>	A3pl	<null>	<null>	A3pl	A3pl	<null>

Table 2. The sentence examples in the proposed dual-source model for English.

English Word Sequence	**And these are fly@@ trap an@@ em@@ ones.**						
Lemma Sequence	and	these	be	fly@@	trap	ane@@	mone	.
POS Tag Sequence	CC	DT	VBP	JJ	JJ	NNS	NNS	.

5. Experiment

5.1. Data Preparation

Following Sennrich *et al.* [26], we merge the WIT corpus [27] that consists of TED talks and the SETimes corpus [28] that consists of news as training corpus, merge dev2010 and tst2010 as validation corpus, and use tst2011, tst2012, tst2013, tst2014 as test corpus. The detailed training and validation corpus statistics of Turkish-English machine translation task are shown in Table 3.

Table 3. The detailed training and validation corpus statistics of Turkish-English.

Dataset	# Sentences	# Turkish Tokens	# English Tokens
Training	355,251	6,712,018	8,376,414
Validation	2,455	39,272	54,061

5.2. Training Parameter

We implemented the proposed dual-source Transformer model by using OpenNMT-tf[3] toolkit. Both the encoder and decoder have 6 layers. The number of hidden units is 512. The number of heads for self-attention is 8. Both the source and target word embedding size are 512, and the number of hidden units in feed-forward layers is 1024. The batch size is 48 sentences. The maximum sentence length is 100. The label smoothing is 0.1. The dropout rate of Transformer is 0.1. The length penalty is 0.6, and the clip gradient is 5.0 [29]. Our parameters are uniformly initialized in [-0.1, 0.1]. We train the model for 120,000 steps by using the Adam optimizer [30] with the learning rate of 0.0002, and we report the result of averaging the 5 last saved checkpoints (saved every 5,000 steps). Decoding is performed by using beam search with the beam size of 5.

We normalize and tokenize the training data, and we use BPE to segment the words and lemmas in Turkish and English by learning separate vocabulary with 32K merge operations. Moreover, we report the case-sensitive BLEU [31] score with tokenization and the ChrF3 [32] score to evaluate the translation performance. The model parameters for Turkish→English (TR-EN) and English→Turkish (EN-TR) machine translation tasks are shown in Table 4.

Table 4. The model parameters for TR-EN and EN-TR machine translation tasks.

Encoder Input		Vocabulary Size		Source Embedding Size	
		Turkish	English	Turkish	English
Word-based Encoder	Words	32,064	31,306	512	512
Knowledge-based Encoder	Lemmas	30,637	30,646	352	384
	POS Tags	14	45	64	128
	Morphological Tags	9,176	-	96	-

5.3. Baseline Models

- NMT baseline

---

[3] https://github.com/OpenNMT/OpenNMT-tf

The Transformer model is employed as our NMT baseline [5]. The model utilizes an encoder to encode the word sequence.

- Single-source model

  We employ the single-source model with linguistic knowledge for comparison [11]. We keep the total embedding size of the linguistic features fixed to 512 as the same with the source word embedding size. More specifically, the embedding sizes of lemma, POS tag, and morphological tag for Turkish are 352, 64, 96, and the embedding sizes of lemma and POS tag for English are 384, 128 as the same with the settings in the proposed dual-source model. In consideration of model efficiency, we only keep the top 50K most frequent lemmas for model training [33].

6. Results

Experimental results of Turkish→English (TR-EN) and English→Turkish (EN-TR) machine translation tasks are shown in Table 5 and Table 6 respectively. For Turkish→English machine translation task, we can see from Table 5 that the proposed dual-source model outperforms both the NMT baseline and the single-source model. It achieves the highest BLEU and ChrF3 scores on all the test datasets, which indicates that the approach is capable of effective improving machine translation quality. Moreover, the dual-source model achieves the highest improvements on tst2014 of 2.4 BLEU points and 1.6 ChrF3 points.

For English→Turkish machine translation task, we can see form Table 6 that the dual-source model achieves the highest BLEU and ChrF3 scores on tst2012, tst2013, tst2014. It achieves the highest improvements on tst2012 of 1.1 BLEU points and 1.5 ChrF3 points. As for the test dataset of tst2011, the proposed model is worse than the NMT baseline on BLEU score, but it is better than the NMT baseline on ChrF3 score. The main reason is that BLEU score is based on the precision of the Turkish words while ChrF3 score is based on both the precision and recall, so the two metrics may occasionally disagree. ChrF3 score was found to correlate well with human judgments, especially for the translation results out of English [34]. Thus, we consider that our model is still better than the NMT baseline on translation performance. Nevertheless, it is not better than the single-source model. The main reason is that the data in tst2011 is not suitable for our dual-source model since its linguistic features of English are not accurate or sufficient enough.

To further evaluate the effectiveness of using various linguistic features in the dual-source model, we separately incorporate single linguistic feature for comparison. Experimental results of using different features for Turkish→English (TR-EN) and English→Turkish (EN-TR) machine translation tasks are shown in Table 7 and Table 8, respectively.

Table 5. The experimental results of TR-EN machine translation task.

Model	BLEU				ChrF3			
	tst2011	tst2012	tst2013	tst2014	tst2011	tst2012	tst2013	tst2014
NMT Baseline	24.18	25.95	26.60	24.98	47.18	48.61	48.65	48.05
Single-Source Model	24.69	26.65	27.43	25.98	47.99	49.26	49.89	48.87
Dual-Source Model	**25.44**	**26.75**	**28.48**	**27.37***	**48.30**	**49.80**	**50.61**	**49.74***

Table 6. The experimental results of EN-TR machine translation task.

Model	BLEU				ChrF3			
	tst2011	tst2012	tst2013	tst2014	tst2011	tst2012	tst2013	tst2014
NMT Baseline	13.82	14.37	13.48	14.73	46.08	46.72	46.15	46.62
Single-Source Model	**13.95**	15.37	14.39	14.90	**46.86**	47.90	47.21	47.50
Dual-Source Model	13.75	**15.48***	**14.44**	**15.36**	46.74	**48.21***	**47.51**	**47.73**

Table 7. The dual-source model with single linguistic feature for TR-EN machine translation task.

Linguistic Feature	BLEU	ChrF3

	tst2011	tst2012	tst2013	tst2014	tst2011	tst2012	tst2013	tst2014
Word	24.18	25.95	26.60	24.98	47.18	48.61	48.65	48.05
+ Lemma	**24.60**	**25.75**	**26.69**	25.20	**47.66**	**48.77**	**49.25**	48.33
+ POS Tag	24.10	25.54	26.20	24.98	47.32	48.66	48.84	47.98
+ Morphological Tag	24.34	25.70	26.39	**25.50**	47.41	48.61	49.08	**48.36**

Table 8. The dual-source model with single linguistic feature for EN-TR machine translation task.

Linguistic Feature	BLEU				ChrF3			
	tst2011	tst2012	tst2013	tst2014	tst2011	tst2012	tst2013	tst2014
Word	13.82	14.37	13.48	14.73	46.08	46.72	46.15	46.62
+ Lemma	13.79	14.59	**13.06**	14.55	45.99	47.11	**46.18**	46.51
+ POS Tag	**14.06**	**15.01**	12.93	**14.93**	**46.52**	**47.98**	45.82	**46.73**

7. Discussion

For Turkish→English machine translation task, we can see from Table 7 that incorporating the lemma feature of Turkish into the proposed dual-source model achieves the highest BLEU and ChrF3 scores on tst2011, tst2012, tst2013 while incorporating morphological tag achieves the highest BLEU and ChrF3 scores on tst2014. For English→Turkish machine translation task, we can see from Table 8 that incorporating the lemma feature of English into the proposed dual-source model achieves the highest BLEU and ChrF3 scores on tst2013 while incorporating POS tag achieves the highest BLEU and ChrF3 scores on tst2011, tst2012, and tst2014. This fact indicates that the external linguistic knowledge is benefit for the conventional NMT model to enhance its representation ability on the source-side language, and various linguistic features are appropriate for different translation tasks and test datasets.

In addition, we also find that incorporating single linguistic feature into the dual-source model sometimes cannot yield improvements on BLEU or ChrF3 scores. The main reason is that the single linguistic feature has not enough effective information for the model training, so the dual-source framework is not conducive to improving the model representation ability of the source-side language. This fact indicates that the approach is capable of better utilizing the external linguistic knowledge and effective integrating them together.

8. Conclusions

This paper proposed a novel dual-source NMT approach that models the words in parallel to the linguistic knowledge by using two separate encoders. The proposed approach is based on the multi-source framework that additionally employs the knowledge-based encoder to specially encode linguistic knowledge. It generalizes the word embedding layer of the knowledge-based encoder to allow for an arbitrary number and type of linguistic features, and it shares all parameters across two encoders to enhance the source-side language representation ability. Moreover, we evaluated the effectiveness of separately incorporating linguistic feature into the proposed dual-source model. We found that the external linguistic knowledge is benefit for the NMT model. Extensive experiments show that the proposed dual-source model achieves significant BLEU and ChrF3 improvements on both Turkish→English and English→Turkish machine translation tasks.

In the future, we plan to employ a more effective combination method of the hidden states and cell states from each encoder to further enhance the dual-source model to learn the representation ability of the source-side language. In addition, we also plan to perform machine translation tasks on the other high-resource language pairs and morphologically-rich languages.

9. Patents

Author Contributions: Investigation, Yirong Pan; Visualization, Yirong Pan; Writing-original draft, Yirong Pan; Writing-review & editing, Xiao Li, Yating Yang and Rui Dong.

References

1. Sutskever, Ilya; Vinyals, Oriol; Le, Quoc V. Sequence to sequence learning with neural networks, Proceedings of the 2014 Conference of the Neural Information Processing Systems, Canada, December 13.

2. Cho, Kyunghyun; Van Merriënboer, Bart; Gulcehre, Caglar; Bahdanau, Dzmitry; Bougares, Fethi; Schwenk, Holger; Bengio, Yoshua. Learning phrase representations using RNN encoder-decoder for statistical machine translation, Proceedings of the 2014 Conference of the Empirical Methods on Natural Language Processing, Qatar, October 25-29.

3. Bahdanau, Dzmitry; Cho, Kyunghyun; Bengio, Yoshua. Neural machine translation by jointly learning to align and translate, Proceedings of the 2015 International Conference on Learning Representations, USA, May 7-9.

4. Luong, Minh-Thang; Pham, Hieu; Manning, Christopher D. Effective approaches to attention-based neural machine translation, Proceedings of the 2015 Conference of the Empirical Methods on Natural Language Processing, Portugal, September 17-21.

5. Vaswani, Ashish; Shazeer, Noam; Parmar, Niki; Uszkoreit, Jakob; Jones, Llion; Gomez, Aidan N; Kaiser, Łukasz; Polosukhin, Illia. Attention is all you need, Proceedings of the 2017 Conference of the Neural Information Processing Systems, UAS, December 4-9.

6. Koehn, Philipp; Och, Franz Josef; Marcu, Daniel. Statistical phrase-based translation, Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology, Canada, May 27-June 1.

7. Shi, Xing; Padhi, Inkit; Knight, Kevin. Does string-based neural MT learn source syntax? Proceedings of the 2016 Conference of the Empirical Methods in Natural Language Processing, USA, November 1-4.

8. Bisazza, Arianna; Federico, Marcello. Morphological pre-processing for Turkish to English statistical machine translation, Proceedings of the 2009 Conference of the International Workshop on Spoken Language Translation, Japan, December 1-2.

9. Ataman, Duygu; Negri, Matteo; Turchi, Marco; Federico, Marcello. Linguistically motivated vocabulary reduction for neural machine translation from Turkish to English. *Prague Bulletin of Mathematical Linguistics* **2017**, *108*, 331-342.

10. Pan Yirong; Li Xiao; Yang Yating; Dong Rui. Morphological Word Segmentation on Agglutinative Languages for Neural Machine Translation. *arXiv preprint arXiv:2001.01589, 2020*.

11. Sennrich, Rico; Haddow, Barry. Linguistic input features improve neural machine translation, Proceedings of the 2016 Conference of the Empirical Methods in Natural Language Processing, Germany, August 11-12.

12. Currey, Anna; Heafield, Kenneth. Multi-source syntactic neural machine translation, Proceedings of the 2018 Conference of the Empirical Methods in Natural Language Processing, Belgium, October 31-November 4.

13. Zoph, Barret; Knight, Kevin. Multi-source neural translation, Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics, USA, June 12-17.

14. Eriguchi, Akiko; Hashimoto, Kazuma; Tsuruoka, Yoshimasa. Tree-to-sequence attentional neural machine translation, Proceedings of the 2016 Conference of the Association for Computational Linguistics, Germany, August 7-12.

15. Yang, Baosong; Wong, Derek F; Xiao, Tong; Chao, Lidia S; Zhu, Jingbo. Towards bidirectional hierarchical representations for attention-based neural machine translation, Proceedings of the 2017 Conference of the Empirical Methods in Natural Language Processing, Denmark, September 9-11.

16. Li, Junhui; Xiong, Deyi; Tu, Zhaopeng; Zhu, Muhua; Zhang, Min; Zhou, Guodong. Modeling source syntax for neural machine translation, Proceedings of the 2017 Conference of the Association for Computational Linguistics, Canada, July 30-August 4.

17. Aqlan, Fares; Fan, Xiaoping; Alqwbani, Abdullah; Al-Mansoub, Akram. Improved Arabic-Chinese machine translation with linguistic input features. *Future Internet* **2019**, *11*, 22.

18. Li, Qiang; Wong, Derek F; Chao, Lidia S; Zhu, Muhua; Xiao, Tong; Zhu, Jingbo; Zhang, Min. Linguistic knowledge-aware neural machine translation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* **2018**, *26*, 2341-2354.

19. Luong, Minh-Thang; Le, Quoc V; Sutskever, Ilya; Vinyals, Oriol; Kaiser, Lukasz. Multi-task sequence to sequence learning, Proceedings of the 2016 International Conference on Learning Representations, Puerto Rico, May 2-4.

20. Junczys-Dowmunt, Marcin; Grundkiewicz, Roman. MS-uedin submission to the wmt2018 ape shared task: Dual-source transformer for automatic post-editing, Proceedings of the 2018 Conference of the Machine Translation, Brussels, October 31-November 1.

21. Liu, Jiangming; Cohen, Shay B; Lapata, Mirella. Discourse representation structure parsing, Proceedings of the 2018 Conference of the Association for Computational Linguistics, Melbourne, July 15-20.

22. Duong, Long; Afshar, Hadi; Estival, Dominique; Pink, Glen; Cohen, Philip R; Johnson, Mark. Multilingual semantic parsing and code-switching, Proceedings of the 2017 Conference of the Computational Natural Language Learning, Canada, August 3-4.

23. van Noord, Rik; Toral, Antonio; Bos, Johan. Linguistic information in neural semantic parsing with multiple encoders, Proceedings of the 2019 International Conference on Computational Semantics, Netherlands, June 3-5.

24. Sak, Haşim; Tunga, Güngör; Murat, Saraçlar. Morphological disambiguation of Turkish text with perceptron algorithm, Proceedings of the 2007 International Conference on Intelligent Text Processing and Computational Linguistics, Berlin.

25. Sennrich, Rico; Haddow, Barry; Birch, Alexandra. Neural machine translation of rare words with subword units, Proceedings of the 2016 Conference of the Association for Computational Linguistics, Germany, August 7-12.

26. Sennrich, Rico; Haddow, Barry; Birch, Alexandra. Improving neural machine translation models with monolingual data, Proceedings of the 2016 Conference of the Association for Computational Linguistics, Germany, August 7-12.

27. Cettolo, Mauro; Girardi, Christian; Federico, Marcello. Wit3: Web inventory of transcribed and translated talks, Proceedings of the 2012 Conference of the European Association for Machine Translation.

28. Tyers, Francis M; Alperen, Murat Serdar. South-east european times: A parallel corpus of balkan languages, Proceedings of the 2010 Conference of the LREC Workshop on Exploitation of Multilingual Resources and Tools for Central and South-Eastern European Languages.

29. Pascanu, Razvan; Mikolov, Tomas; Bengio, Yoshua. On the difficulty of training recurrent neural networks, Proceedings of the 2013 International Conference on Machine Learning, USA, June 16-21.

30. Kingma, Diederik P; Ba, Jimmy. Adam: A method for stochastic optimization, Proceedings of the 2015 International Conference on Learning Representations, USA, May 7-9.

31. Papineni, Kishore; Roukos, Salim; Ward, Todd; Zhu, Wei-Jing. BLEU: A method for automatic evaluation of machine translation, Proceedings of the 2002 Conference of the Association for Computational Linguistics, USA, July 6-12.

32. Popović, Maja. ChrF: Character n-gram F-score for automatic mt evaluation. Proceedings of the 2015 Workshop on Statistical Machine Translation, Portugal, September 17-18.

33. Jean, Sébastien; Cho, Kyunghyun; Memisevic, Roland; Bengio, Yoshua. On using very large target vocabulary for neural machine translation, Proceedings of the 2015 Conference of the Association for Computational Linguistics, China, July 26-31.

34. Stanojević, Miloš; Kamran, Amir; Koehn, Philipp; Bojar Ondřej. Results of the wmt15 metrics shared task. Proceedings of the 2015 Workshop on Statistical Machine Translation, Portugal, September 17-18.