

# Syntax in End-to-End Natural Language Processing

Hai Zhao<sup>1</sup>, Rui Wang<sup>1</sup>, and Kehai Chen<sup>2</sup>

<sup>1</sup> Department of Computer Science and Engineering,  
Shanghai Jiao Tong University, Shanghai, China

<sup>2</sup> Advanced Translation Technology Laboratory,

National Institute of Information and Communications Technology, Kyoto, Japan

zhaohai@cs.sjtu.edu.cn, wangrui.nlp@gmail.com, khchen@nict.go.jp

## Abstract

This tutorial surveys the latest technical progress of syntactic parsing and the role of syntax in end-to-end natural language processing (NLP) tasks, in which semantic role labeling (SRL) and machine translation (MT) are the representative NLP tasks that have always been beneficial from informative syntactic clues since a long time ago, though the advance from end-to-end deep learning models shows new results. In this tutorial, we will first introduce the background and the latest progress of syntactic parsing and SRL/NMT. Then, we will summarize the key evidence about the syntactic impacts over these two concerning tasks, and explore the behind reasons from both computational and linguistic background.

## 1 Tutorial Content

**Syntax** is the insightfulness about formal relative position inside languages, whose mathematical formalism was pioneered by [Chomsky \(1957\)](#). Syntactic parsing has been enduring for a significant progress since deep learning was fully introduced into natural language processing (NLP). We identify two development stages for parsing techniques by considering whether deep learning was involved or not. For the parsers that were built on traditional machine learning models, most work focus on designing better search algorithms or better structural modeling about syntax, while few ever consider feature engineering. For the parsers using deep learning models, most work turn to more effective and more salient representations, following the same structural formalization since the times of traditional parsers. We observe a series of significant performance improvement since 2014 ([Chen and Manning, 2014](#); [Dozat and Manning, 2017](#)). In this part, we will survey

the key language representation improvement for syntactic parsing. In general, syntactic information contributes to other end-to-end NLP tasks, such as SRL and MT. We summarize the contribution of syntax to SRL and MT in [Table 1](#). **Syntax in SRL.** SRL or semantic parsing as a computational job started since different semantic annotated datasets were released in recent two decades, which is trained by using PropBank such as [Palmer et al. \(2005\)](#). During treebank annotation, the semantic annotation may be naturally assigned onto syntactic constituents, so that it makes sense that the latter may help the former in either of linguistic explanation or machine learning procedure. Considering syntactic information helps or not, the performance variation of SRL may range about 5-10% in terms of traditional models. However, there has come new results since end-to-end SRL was proposed. Nearly all state-of-the-art SRL models, either span or dependency, have been based on LSTM backbone since [Zhou and Xu \(2015a\)](#). We attribute such a change of syntactic role to the effective distributional and contextualized representation offered by the LSTM from word embedding. Note that word embedding may have both syntactic and semantic sense.

Since the method by [Zhou and Xu \(2015b\)](#) and [Marcheggiani et al. \(2017\)](#), deep-learning-based SRL has obtained much less contribution from syntactic input. For either span or dependency SRL, deep models receive a less than 2% performance improvement even when perfect syntax (gold syntax labels) is introduced as shown by [He et al. \(2017a\)](#) and [He et al. \(2018a\)](#). We re-implemented the model of [Li et al. \(2019\)](#) and introduced a syntactic constraint in their span selection from a strong parser, which indicates that stronger syntax-agnostic models receive less enhancement from syntax information.

Tasks	Attention Mechanism			PreLM	Syntax	Effectiveness
	attention	self-attention	biaffine			
Syntactic parsing			++	++		++
SRL		++				++
			++	++	+	++
NMT	RNN		++	0	+	++
	Self-attention			0	-	-

Table 1: Role of different technical factors for the three NLP tasks. “++” denotes the significant performance contribution when used alone; “+” denotes the moderate contribution; “0” denotes mainly studies in zero/low-resource scenarios; “-” denotes negative or little impact. The mark in the rightmost column indicates whether it is overall effective when all marked factors to the left are combined.

**Syntax in MT** also endures a methodology change from statistical machine translation (SMT) (Brown et al., 1993) to neural machine translation (NMT) (Sutskever et al., 2014; Bahdanau et al., 2015) as the task of SRL. For typical SMT, besides phrase based SMT (Och et al., 1999; Koehn et al., 2003), syntactic (tree) based methods have been well developed (Yamada and Knight, 2001; Mi et al., 2008). In some scenarios, especially when the domain of the MT corpus is similar to the domain of the parsing corpus, the performance of tree based SMT is better than phrase based SMT (Koehn, 2009). For NMT, it so far achieves significant progress by using end-to-end based structure since 2014 (Sutskever et al., 2014; Bahdanau et al., 2015). Recently, self-attention based transformer (Vaswani et al., 2017) has become new state-of-the-art architecture in NMT and gives a series of new state-of-the-art benchmarks (Bojar et al., 2018; Marie et al., 2018; Wang et al., 2018a; Marie et al., 2019). Syntax information has been shown that it can improve the performances of the recurrent neural network (RNN) based NMT on conditions (Eriguchi et al., 2016, 2017; Chen et al., 2017a; Li et al., 2017; Wu et al., 2017; Chen et al., 2017b, 2018). However, so far it has not been shown significantly widely useful in self-attention based NMT. There are only a few work (Ma et al., 2019) adopted the syntactic information into the positional embedding of Transformer. We will give a detailed analysis on this issue by surveying the key technique details.

**Linguistic in MT.** In addition, we will investigate why linguistic cognition and prior knowledge can enhance the control of the dominant end-to-end neural framework, which makes the translation between a language pair proceed according to the expected and interpretable way. On one hand, linguistic cognition enables translation model (1) to reduce translation errors that violate

common sense, such as over/under-translation questions (Tu et al., 2016), troublesome words modeling (Zhao et al., 2018b) and so on; (2) to have some basic abilities of human translator, for example, word importance modeling (Chen et al., 2020), translation refinement (Song et al., 2020), structured information (Xu et al., 2020), diverse feature (Chen et al., 2020) and so on. On the other hand, linguistic prior knowledge (i.e. alignment, bilingual lexicon, phrase table, and knowledge graphs) to alleviate the problem of inadequacy target translations which are caused by the language model property of the encoder-decoder framework (Feng et al., 2017; Zhang et al., 2017; Zhao et al., 2018a; Wang et al., 2018b). Moreover, linguistic differences between the source language and target language can learn natural language representations that are easy to be understood by the translation model, for example, word order difference (Chen et al., 2019; Ding et al., 2020), morphological differences (Ji et al., 2019) and so on. Meanwhile, linguistic shared feature between the source language and target language can also enhance the understanding and generation of natural language in MT, for example, shared words (Artetxe et al., 2018), image information (Yin et al., 2020), video information (Wang et al., 2020) and so on.

## 2 Relevance to the Computational Linguistics Community

The topics included in this tutorial, i.e., syntax parsing, SRL, and MT, are all the classic ones to the entire NLP/CL community. This tutorial is primarily towards researchers who have a basic understanding of deep learning based NLP. We believe that this tutorial would help the audience more deeply understand the relationship between three classic NLP tasks, i.e., syntax parsing and SRL/MT.

Presenter: Hai Zhao	Presenter: Rui Wang and Kehai Chen	
1. Syntactic Parsing (50 min)	3. Syntax in MT (40 min)	4. Summary (20 min)
1.1 Traditional syntactic parsing	3.1 Basics of MT	4.1 Conclusion
1.2 Neural syntactic parsing	3.2 Syntax in RNN-based MT	4.2 Future trends
1.3 Basic of end-to-end NLP	3.3 Syntax in self-attention based MT	
2. Syntax in SRL (40 min)	4.Linguistic in MT (30 min)	
2.1 Basic of SRL	4.1 Linguistic cognition for MT	
2.2 Linguistic, Syntax, and Semantics	4.2 Linguistic prior knowledge for MT	
2.3 Syntax in end-to-end base SRL		
Coffee Break (30 min)		

Table 2: Tutorial outlines

### 3 Type of the Tutorial: Cutting-edge

We introduce the cutting-edge technologies.

### 4 Tutorial Outlines

We will present our tutorial in three hours. The detailed tutorial outlines are shown in Table 1.

### 5 Breadth

20-30% of the tutorial covers work by the tutorial presenters and 70-80% by other researchers.

### 6 Diversity Considerations

N/A

### 7 Specification of Any Prerequisites for the Attendees

This tutorial is primarily aimed at researchers who have a basic understanding of NLP.

### 8 Small reading list

- Deep Learning: *Deep learning* (LeCun et al., 2015)
- Syntactic Parsing: *Deep biaffine attention for neural dependency parsing* (Dozat and Manning, 2016) and *Constituency parsing with a self-attentive encoder* (Kitaev and Klein, 2018).
- SRL: *Syntax for semantic role labeling, to be, or not to be* (He et al., 2018b) and *Deep semantic role labeling: What works and whats next* (He et al., 2017b).
- Machine Translation: *Statistical machine translation* (Koehn, 2009) and *Neural machine translation by jointly learning to align and translate* (Bahdanau et al., 2015).

### 9 Presenters

1. Dr. Hai Zhao, Professor, Department of Computer Science and Engineering, Shanghai Jiao Tong University, China.

[zhaohai@cs.sjtu.edu.cn](mailto:zhaohai@cs.sjtu.edu.cn)

<http://bcmi.sjtu.edu.cn/~zhaohai>

His research interest is natural language processing. He has published more than 120 papers in ACL, EMNLP, COLING, ICLR, AAAI, IJCAI, and IEEE TKDE/TASLP. He won the first places in several NLP shared tasks, such as CoNLL and SIGHAN Bakeoff and top ranking in remarkable machine reading comprehension task leaderboards such as SQuAD2.0 and RACE.

He has taught the course “natural language processing” in SJTU for more than 10 years. He is ACL-2017 area chair on parsing, and ACL-2018/2019 (senior) area chairs on morphology and word segmentation.

2. Dr. Rui Wang, Tenured Researcher, Advanced Translation Technology Laboratory, National Institute of Information and Communications Technology (NICT), Japan

[wangrui.nlp@gmail.com](mailto:wangrui.nlp@gmail.com)

<https://wangruinlp.github.io>

His research focuses on machine translation (MT), a classic task in NLP. His recent interests are traditional linguistic based and cutting-edge machine learning based approaches for MT. He (as the first or the corresponding authors) has published more than 30 MT papers in top-tier NLP/ML/AI conferences and journals, such as ACL, EMNLP, ICLR, AAAI, IJCAI, IEEE/ACM transactions, etc. He has also won several first places in top-tier MT shared tasks, such as WMT-2018, WMT-2019, WMT-2020, etc.

He has given several tutorial and invited talks in

conferences, such as CWMT, CCL, etc. He served as the area chairs of ICLR-2021 and NAACL-2021.

3. Dr. Kehai Chen, Postdoctoral Researcher, Advanced Translation Technology Laboratory, National Institute of Information and Communications Technology (NICT), Japan

[khchen@nict.go.jp](mailto:khchen@nict.go.jp)

<https://chenkehai.github.io>

His research focuses on linguistic-motivated machine translation (MT), a classic NLP task in AI. He has published more than 20 MT and NLP papers in top-tier NLP/ML/AI conferences and journals, such as ACL, ICLR, AAAI, EMNLP, IEEE/ACM Transactions on Audio, Speech, and Language Processing, ACM Transactions on Asian and Low-Resource Language Information Processing, etc. He served as a senior program committee of AAAI-2021.

## 10 Previous Venues and Approximate Audience Sizes

There are some tutorials focusing on single NLP tasks, such as NMT in ACL-2016/IJCNLP-2018, semantic parsing in ACL-2018. In particular, the NMT tutorial at ACL-2016 (with around 800 registrations) had attracted around 150 attendees and the one at IJCNLP-2017 (with around 300 registrations) had attracted around 40 attendees.

Our tutorial will become the first one that explores the relationship between syntactic impact and end-to-end NLP tasks. As our topic is rather broader, we hope that this tutorial will attract around 100-200 attendees.

## 11 Special Requirements

None

## 12 Preferable Venue(s)

ACL-IJCNLP/EMNLP/NAACL-HLT/EACL

## 13 Open Access

Yes

## References

Mikel Artetxe, Gorra Labaka, Eneko Agirre, and Kyunghyun Cho. 2018. [Unsupervised neural machine translation](#). In *International Conference on Learning Representations*.

Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2015. [Neural machine translation by jointly learning to align and translate](#). In *International Conference on Learning Representations*, San Diego, CA.

Ondej Bojar, Rajen Chatterjee, Christian Federmann, Mark Fishel, Yvette Graham, Barry Haddow, Matthias Huck, Antonio Jimeno Yepes, Philipp Koehn, Christof Monz, Matteo Negri, Aurlie Nvol, Mariana Neves, Matt Post, Lucia Specia, Marco Turchi, and Karin Verspoor. 2018. [Findings of the 2018 conference on machine translation \(WMT18\)](#). In *Proceedings of the Third Conference on Machine Translation*, pages 272–307, Belgium, Brussels.

Peter F. Brown, Vincent J. Della Pietra, Stephen A. Della Pietra, and Robert L. Mercer. 1993. [The mathematics of statistical machine translation: Parameter estimation](#). *Computational Linguistics*, 19(2):263–311.

Danqi Chen and Christopher D. Manning. 2014. A fast and accurate dependency parser using neural networks. In *Proceedings of EMNLP 2014*.

Huadong Chen, Shujian Huang, David Chiang, and Jiajun Chen. 2017a. [Improved neural machine translation with a syntax-aware encoder and decoder](#). In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1936–1945, Vancouver, Canada.

K. Chen, R. Wang, M. Utiyama, E. Sumita, T. Zhao, M. Yang, and H. Zhao. 2020. [Towards more diverse input representation for neural machine translation](#). *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 28:1586–1597.

Kehai Chen, Rui Wang, Masao Utiyama, Lemao Liu, Akihiro Tamura, Eiichiro Sumita, and Tiejun Zhao. 2017b. [Neural machine translation with source dependency representation](#). In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 2846–2852, Copenhagen, Denmark.

Kehai Chen, Rui Wang, Masao Utiyama, and Eiichiro Sumita. 2019. [Neural machine translation with reordering embeddings](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 1787–1799, Florence, Italy. Association for Computational Linguistics.

Kehai Chen, Rui Wang, Masao Utiyama, and Eiichiro Sumita. 2020. [Content word aware neural machine translation](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 358–364, Online. Association for Computational Linguistics.

Kehai Chen, Rui Wang, Masao Utiyama, Eiichiro Sumita, and Tiejun Zhao. 2018. [Syntax-directed](#)

- attention for neural machine translation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 4792–4799, New Orleans, LA.
- Noam Chomsky. 1957. *Syntactic structures*. Mouton.
- Liang Ding, Longyue Wang, and Dacheng Tao. 2020. Self-attention with cross-lingual position representation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1679–1685, Online. Association for Computational Linguistics.
- Timothy Dozat and Christopher D Manning. 2016. Deep biaffine attention for neural dependency parsing. *arXiv preprint arXiv:1611.01734*.
- Timothy Dozat and Christopher D. Manning. 2017. Deep biaffine attention for neural dependency parsing. In *Proceedings of ICLR*.
- Akiko Eriguchi, Kazuma Hashimoto, and Yoshimasa Tsuruoka. 2016. Tree-to-sequence attentional neural machine translation. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 823–833, Berlin, Germany.
- Akiko Eriguchi, Yoshimasa Tsuruoka, and Kyunghyun Cho. 2017. Learning to parse and translate improves neural machine translation. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 72–78, Vancouver, Canada.
- Yang Feng, Shiyue Zhang, Andi Zhang, Dong Wang, and Andrew Abel. 2017. Memory-augmented neural machine translation. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pages 1390–1399, Copenhagen, Denmark. Association for Computational Linguistics.
- Luheng He, Kenton Lee, Mike Lewis, and Luke Zettlemoyer. 2017a. Deep semantic role labeling: What works and what’s next. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 473–483, Vancouver, Canada.
- Luheng He, Kenton Lee, Mike Lewis, and Luke Zettlemoyer. 2017b. Deep semantic role labeling: What works and what’s next. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 473–483, Vancouver, Canada. Association for Computational Linguistics.
- Shexia He, Zuchao Li, Hai Zhao, and Hongxiao Bai. 2018a. Syntax for semantic role labeling, to be, or not to be. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2061–2071, Melbourne, Australia.
- Shexia He, Zuchao Li, Hai Zhao, and Hongxiao Bai. 2018b. Syntax for semantic role labeling, to be, or not to be. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2061–2071, Melbourne, Australia. Association for Computational Linguistics.
- Yatu Ji, Hongxu Hou, Chen Junjie, and Nier Wu. 2019. Improving Mongolian-Chinese neural machine translation with morphological noise. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, pages 123–129, Florence, Italy. Association for Computational Linguistics.
- Nikita Kitaev and Dan Klein. 2018. Constituency parsing with a self-attentive encoder. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2676–2686, Melbourne, Australia.
- Philipp Koehn. 2009. *Statistical machine translation*. Cambridge University Press.
- Philipp Koehn, Franz Josef Och, and Daniel Marcu. 2003. Statistical phrase-based translation. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology - Volume 1*, pages 48–54, Edmonton, Canada.
- Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. 2015. Deep learning. *nature*, 521(7553):436.
- Junhui Li, Deyi Xiong, Zhaopeng Tu, Muhua Zhu, Min Zhang, and Guodong Zhou. 2017. Modeling source syntax for neural machine translation. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 688–697, Vancouver, Canada.
- Zuchao Li, Shexia He, Hai Zhao, Yiqing Zhang, Zhuosheng Zhang, Xi Zhou, and Xiang Zhou. 2019. Dependency or span, end-to-end uniform semantic role labeling. In *The Thirty-Third AAAI Conference on Artificial Intelligence*, Honolulu, Hawaii.
- Chunpeng Ma, Akihiro Tamura, Masao Utiyama, Eiichiro Sumita, and Tiejun Zhao. 2019. Improving neural machine translation with neural syntactic distance. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 2032–2037, Minneapolis, Minnesota. Association for Computational Linguistics.
- Diego Marcheggiani, Anton Frolov, and Ivan Titov. 2017. A simple and accurate syntax-agnostic neural model for dependency-based semantic role labeling. In *Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017)*, pages 411–420, Vancouver, Canada.

- Benjamin Marie, Haipeng Sun, Rui Wang, Kehai Chen, Atsushi Fujita, Masao Utiyama, and Eiichiro Sumita. 2019. [NICT’s unsupervised neural and statistical machine translation systems for the WMT19 news translation task](#). In *Proceedings of the Fourth Conference on Machine Translation (Volume 2: Shared Task Papers, Day 1)*, pages 294–301, Florence, Italy. Association for Computational Linguistics.
- Benjamin Marie, Rui Wang, Atsushi Fujita, Masao Utiyama, and Eiichiro Sumita. 2018. [NICT’s neural and statistical machine translation systems for the WMT18 news translation task](#). In *Proceedings of the Third Conference on Machine Translation, Volume 2: Shared Task Papers*, pages 453–459, Brussels, Belgium.
- Haitao Mi, Liang Huang, and Qun Liu. 2008. [Forest-based translation](#). In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, pages 192–199, Columbus, Ohio.
- Franz Josef Och, Christoph Tillmann, and Hermann Ney. 1999. [Improved alignment models for statistical machine translation](#). In *Proceedings of the 1999 Joint SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora*, College Park, MD.
- Martha Palmer, Daniel Gildea, and Paul Kingsbury. 2005. [The proposition bank: An annotated corpus of semantic roles](#). *Computational Linguistics*, 31(1):71–106.
- Kaitao Song, Xu Tan, and Jianfeng Lu. 2020. [Neural machine translation with error correction](#). In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20*, pages 3891–3897. International Joint Conferences on Artificial Intelligence Organization. Main track.
- Ilya Sutskever, Oriol Vinyals, and Quoc VV Le. 2014. [Sequence to sequence learning with neural networks](#). In *Advances in Neural Information Processing Systems*, pages 3104–3112, Montreal, Canada.
- Zhaopeng Tu, Zhengdong Lu, Yang Liu, Xiaohua Liu, and Hang Li. 2016. [Modeling coverage for neural machine translation](#). In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 76–85, Berlin, Germany. Association for Computational Linguistics.
- Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. [Attention is all you need](#). *CoRR*, abs/1706.03762.
- Rui Wang, Chenchen Ding, Masao Utiyama, and Eiichiro Sumita. 2018a. [English-Myanmar NMT and SMT with pre-ordering: NICT’s machine translation systems at WAT-2018](#). In *The 5th Workshop on Asian Translation*, Hong Kong, China.
- X. Wang, Z. Tu, and M. Zhang. 2018b. [Incorporating statistical machine translation word knowledge into neural machine translation](#). *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 26(12):2255–2266.
- Xin Wang, Jesse Thomason, Ronghang Hu, Xinlei Chen, Peter Anderson, Qi Wu, Asli Celikyilmaz, Jason Baldridge, and William Yang Wang, editors. 2020. [Proceedings of the First Workshop on Advances in Language and Vision Research](#). Association for Computational Linguistics, Online.
- Shuangzhi Wu, Dongdong Zhang, Nan Yang, Mu Li, and Ming Zhou. 2017. [Sequence-to-dependency neural machine translation](#). In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 698–707, Vancouver, Canada.
- Hongfei Xu, Josef van Genabith, Deyi Xiong, Qihui Liu, and Jingyi Zhang. 2020. [Learning source phrase representations for neural machine translation](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 386–396, Online. Association for Computational Linguistics.
- Kenji Yamada and Kevin Knight. 2001. [A syntax-based statistical translation model](#). In *Proceedings of the 39th Annual Meeting on Association for Computational Linguistics*, pages 523–530, Toulouse, France.
- Yongjing Yin, Fandong Meng, Jinsong Su, Chulun Zhou, Zhengyuan Yang, Jie Zhou, and Jiebo Luo. 2020. [A novel graph-based multi-modal fusion encoder for neural machine translation](#). In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 3025–3035, Online. Association for Computational Linguistics.
- Jiacheng Zhang, Yang Liu, Huanbo Luan, Jingfang Xu, and Maosong Sun. 2017. [Prior knowledge integration for neural machine translation using posterior regularization](#). In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1514–1523, Vancouver, Canada. Association for Computational Linguistics.
- Yang Zhao, Yining Wang, Jiajun Zhang, and Chengqing Zong. 2018a. [Phrase table as recommendation memory for neural machine translation](#). In *Proceedings of the 27th International Joint Conference on Artificial Intelligence, IJCAI’18*, page 46094615. AAAI Press.
- Yang Zhao, Jiajun Zhang, Zhongjun He, Chengqing Zong, and Hua Wu. 2018b. [Addressing troublesome words in neural machine translation](#). In *Proceedings*

of the 2018 Conference on Empirical Methods in Natural Language Processing, pages 391–400, Brussels, Belgium. Association for Computational Linguistics.

Jie Zhou and Wei Xu. 2015a. [End-to-end learning of semantic role labeling using recurrent neural networks](#). In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1127–1137, Beijing, China.

Jie Zhou and Wei Xu. 2015b. [End-to-end learning of semantic role labeling using recurrent neural networks](#). In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1127–1137, Beijing, China.