Neural Sentence Simplification with Semantic Dependency Information

Zhe Lin And Xiaojun Wan

CONTENTS

01. Introduction

02. Model

03. Experiments

04. Conclusion

PART 01

Introduction

Introduction

Sentence simplification, which is a type of paraphrase task, aims to reduce the linguistic complexity of a sentence, while still preserving its salient information and meaning.

Sentence simplification has many practical applications. For instance, it can provide assistance for low-literacy reader (Watanabe et al. 2009) or for patients with linguistic and cognitive disabilities (Carroll et al. 1999). In addition, a simplification component could be used to improve the performance of tasks such as parsing (Chandrasekar, Doran, and Srinivas 1996), summarization (Klebanov, Knight, and Marcu 2004) and so on.

Introduction

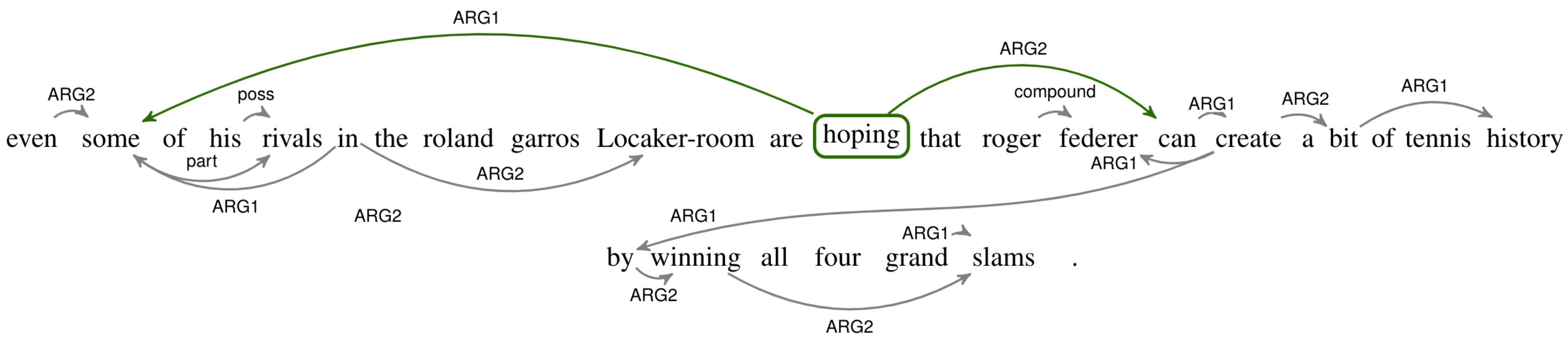
An example for sentence simplification using different models.

Source	To prevent overfishing , the agreement would , among other things , make it much easier to establish marine protected areas -LRB- MPAs -RRB- in the high seas .
Reference	An agreement would make it easier to create marine protected areas .
Transformer	To prevent lionfish , the agreement would among other things , make much easier to protected areas -LRB- MPAs -RRB- in the high seas .
SDISS (ours)	The agreement would make it much easier to establish marine protected areas .

Introduction

Many researches in other tasks have got positive effect by introducing semantic information into neural models. But few attempts have been made for the task of sentence simplification yet.

In this study, we aim to investigate the use of semantic information in neural text simplification systems and we focus on the semantic dependency graph of the source sentence.



PART 02

Model

Overview of model

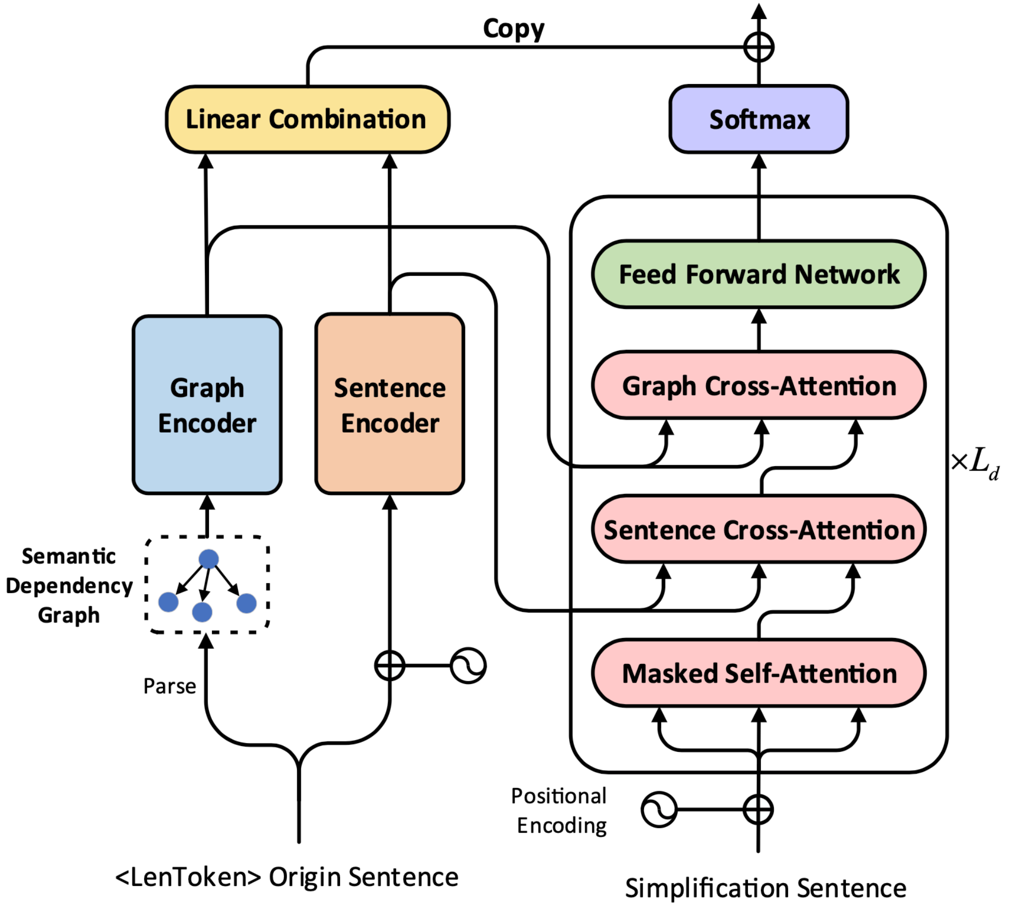


Figure 1: An overview of SDISS, which consists of sentence encoder, graph encoder and sentence decoder.

Graph Construction

In order to improve the information propagation process in the graph, we add reverse edge to represent the relationship from tail node to head node. Then, we split the directed graph into a forward graph and a reversed graph.

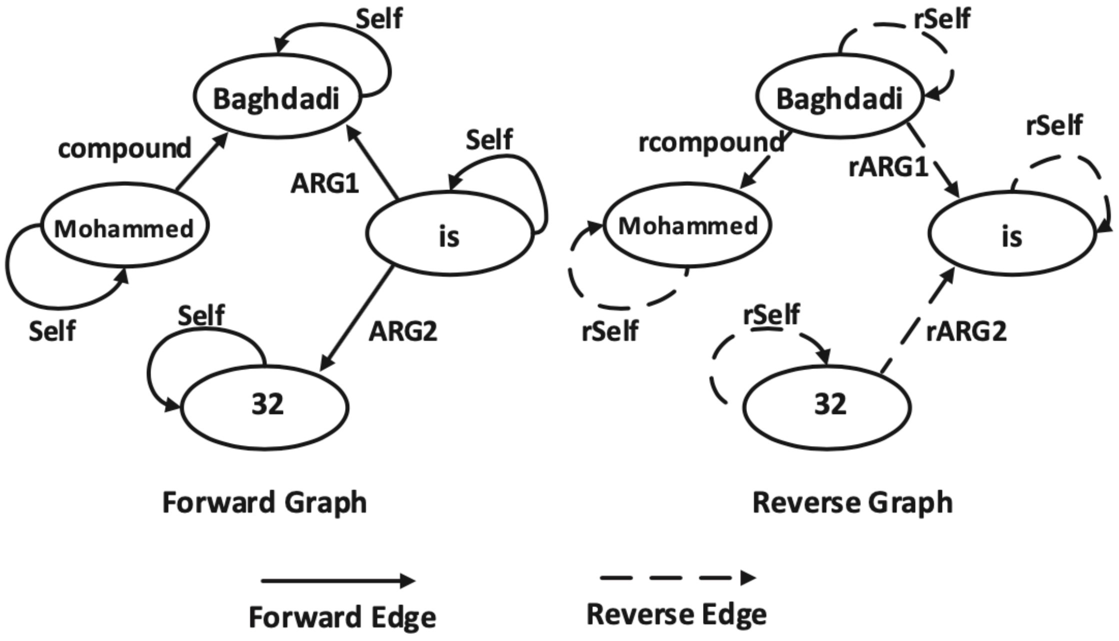


Figure 2: Example forward and reverse graphs for the sentence 'Mohammed Baghdadi is 32.'

Graph Encoder

Graph Node information Aggregation

$$
\begin{aligned}
v_{agg} &= [v_{tail} \| e_{type}]\mathbf{W}_{agg} + \mathbf{b}_{agg} \\
gate &= \sigma([v_{tail} \| e_{type}]\mathbf{W}_{gate} + \mathbf{b}_{gate}) \\
v &= gate \otimes v_{agg}
\end{aligned}
$$

This equation means that we first aggregate tail node information with edge information, and then extract the useful information by a gate unit, which can be seen as a method of pruning information.

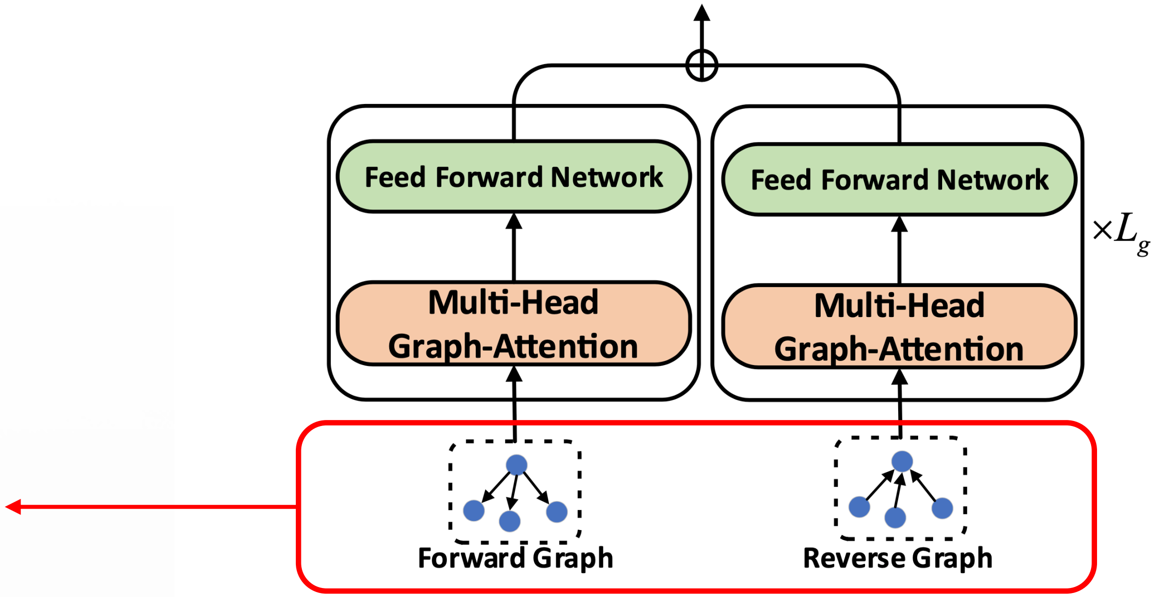


Figure 3: The structure of our graph encoder. The whole graph encoder consists of two component graph encoders, one for forward graph and another for reverse graph. Each component graph encoder is a stack of L_g identical layers, with a multi-head graph-attention block and a feed forward network. Finally, we combine the outputs of two encoders as the output of graph encoder.

北京大學 PEKING UNIVERSITY

Graph Encoder

Multi-Head Graph Attention

$$\alpha_{uv} = \frac{\exp\left(\phi\left(\mathbf{a}^\top\left[\mathbf{W}_a u \| \mathbf{W}_a v\right]\right)\right)}{\sum_{v_k \in \mathcal{N}_u} \exp\left(\phi\left(\mathbf{a}^\top\left[\mathbf{W}_a u \| \mathbf{W}_a v_k\right]\right)\right)}$$

$$\mathrm{GAT}(u) = \sum_{v \in \mathcal{N}_u} \alpha_{uv} v$$

$$\begin{aligned}
\hat{g}_i^j &= \mathrm{GAT}(g_i \mathbf{W}_h^j) \\
\mathrm{Head}_j &= (\hat{g}_1^j, \hat{g}_2^j, \cdots, \hat{g}_N^j) \\
\mathrm{MultiHeadGAT}(g) &= \left(\Big\|_{j=1}^{H} \mathrm{Head}_j \right) \mathbf{W}_o
\end{aligned}$$

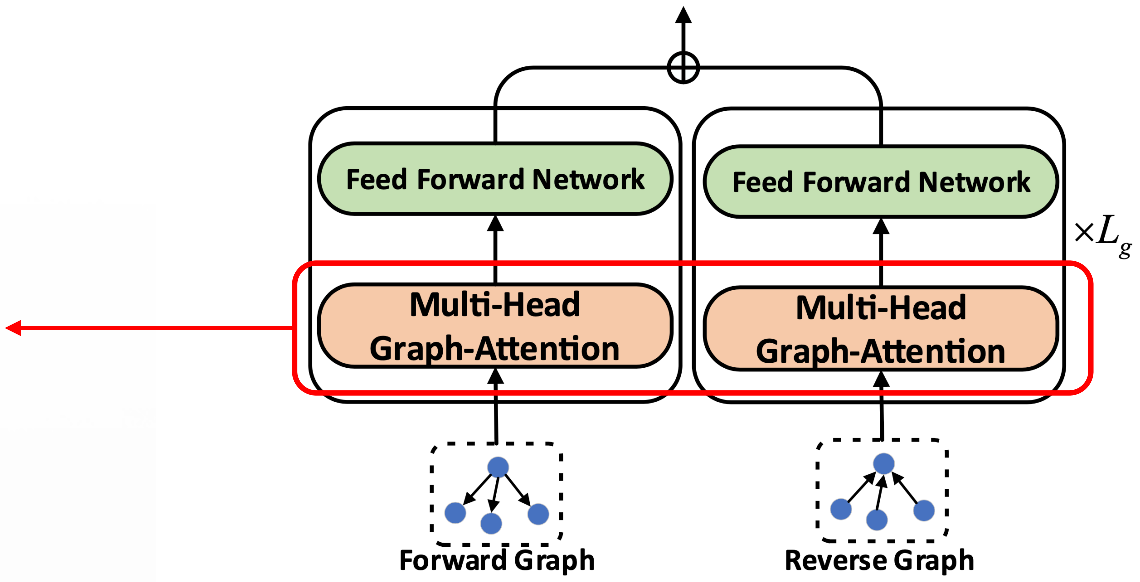


Figure 3: The structure of our graph encoder. The whole graph encoder consists of two component graph encoders, one for forward graph and another for reverse graph. Each component graph encoder is a stack of L_g identical layers, with a multi-head graph-attention block and a feed forward network. Finally, we combine the outputs of two encoders as the output of graph encoder.

Graph Encoder

Feed Forward Network

$$\tilde{g}^l = \mathrm{LN}(g^{l-1} + \mathrm{MultiHeadGAT}(g^{l-1}))$$
$$g^l = \mathrm{LN}(\tilde{g}^l + \mathrm{FFN}(\tilde{g}^l))$$

$$g = \overrightarrow{g} + \overleftarrow{g}$$

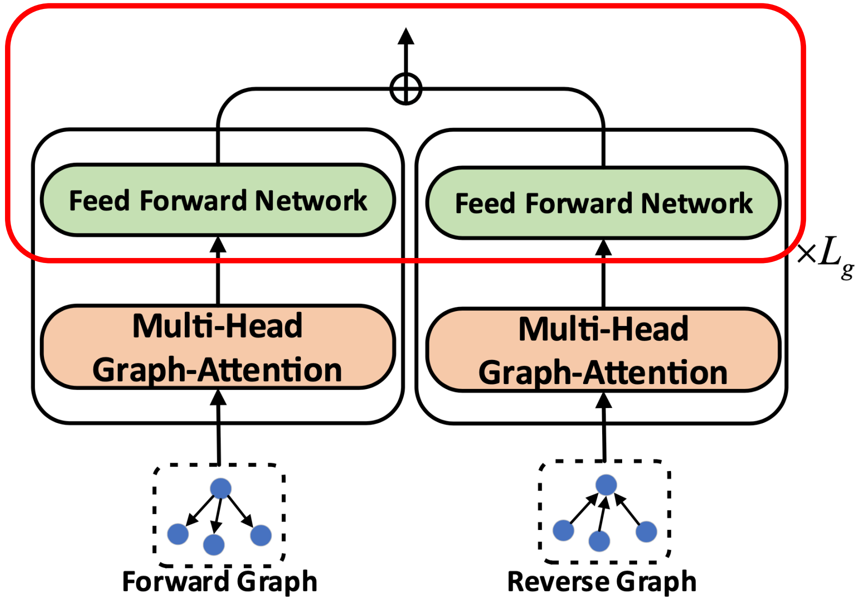


Figure 3: The structure of our graph encoder. The whole graph encoder consists of two component graph encoders, one for forward graph and another for reverse graph. Each component graph encoder is a stack of L_g identical layers, with a multi-head graph-attention block and a feed forward network. Finally, we combine the outputs of two encoders as the output of graph encoder.

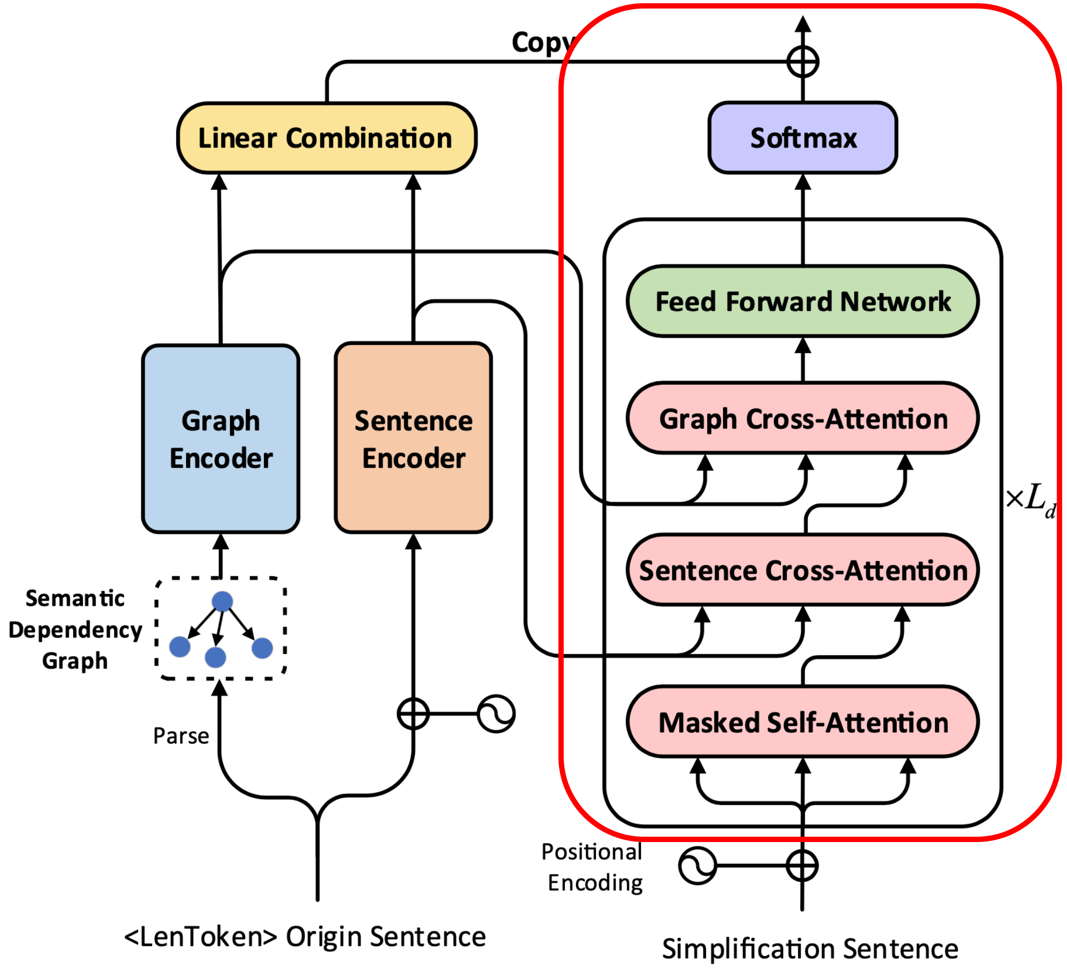
Sentence Decoder

$$\tilde{d}^l = \mathrm{LN}(d^{l-1} + \mathrm{MultiHeadAtt}(d^{l-1}, d^{l-1}, d^{l-1}))$$
$$\tilde{d}^l_s = \mathrm{LN}(\tilde{d}^l + \mathrm{MultiHeadAtt}(\tilde{d}^l, s, s))$$
$$\tilde{d}^l_g = \mathrm{LN}(\tilde{d}^l_s + \mathrm{MultiHeadAtt}(\tilde{d}^l_s, g, g))$$
$$d^l = \mathrm{LN}(\tilde{d}^l_g + \mathrm{FFN}(\tilde{d}^l_g))$$
$$P_g = \mathrm{softmax}(d^{L_d}\mathbf{W}_{out} + \mathbf{b}_{out})$$

Copy Mechanism

$$h_{comb} = \mathrm{ReLU}([g \| s]\mathbf{W}_{comb} + \mathbf{b}_{comb})$$
$$P_{copy} = \mathrm{softmax}(d^{L_d} h^T_{comb} + \mathbf{b}_\epsilon)$$

$$\eta = \sigma(d^{L_d}\mathbf{W}_{eta} + \mathbf{b}_{eta})$$
$$P = \eta P_g + (1-\eta)P_{copy}$$



Copy Loss

We find that copy mechanism tends to copy original sentence completely without restriction. So we raise copy loss to penalize over-replication. Intuitively, adding the probability of copying from the original sentence into the loss can make copy mechanism avoid over-replication. So our loss function is as follows:

$$loss_c = \sum_{i=1}^{d_s} (1-\eta) p_c^i(w_t)$$

$$loss(w_t) = -log\, P(w_t) + \lambda \times loss_c$$

Length Token

We employ LenTok to compress outputs. We add **<SHORT>** , **<MIDDLE>** and **<LONG>** tokens to the original sentence according to its target sentence's length in the training set.

<LenTok> Source Sentence $\longrightarrow$ Simplified Sentence

If the target sentence's length is less than t_{min}, we add **<SHORT>** to the beginning of the original sentence, and **<LONG>** will be added if the target sentence's length is greater than t_{max}. The other sentences with target sentence's length during $[t_{min}, t_{max}]$ are added with **<MIDDLE>.** During testing, we add **<MIDDLE>** to all original sentences. This can help to improve sentence quality.

PART 03

Experiments

Experiments

- Sentence Simplification Datasets
  - Newsela
  - WikiSmall
  - WikiLarge
    - with 8 references writed by human
- Evaluation Metrics
  - SARI
  - FKGL
  - Unchanged Percentage(unc%)
  - BLEU is not suitable for this task.

Dataset	Vocab Size src	Vocab Size tgt	token/sent src	token/sent tgt
Newsela	41,066	30,193	25.94	15.89
WikiSmall	113,368	93,835	24.26	20.33
WikiLarge	201,841	168,962	25.17	18.51

Results

WikiLarge	BLEU	SARI ↑	FKGL ↓	% unc.
Reference	-	-	8.88	15.88
PBMT-R	81.81	38.56	8.33	**10.58**
Hybrid	48.97	31.40	**4.57**	36.21
Transformer	84.27	35.34	7.92	23.08
DRESS	77.18	37.08	6.59	22.28
DRESS-Ls	80.12	37.27	6.62	27.02
NSELSTM-S	80.43	36.88	-	-
NSELSTM-B	92.02	33.43	-	-
EditNTS	86.68	38.22	7.30	10.86
SDISS(ours)	77.36	**38.66**	7.07	13.37
Models with external human knowledge				
SBMT-SARI	73.08	39.96	7.29	9.47
DMASS+DCSS	80.53	40.45	7.79	6.69

WikiSmall	BLEU	SARI ↑	FKGL ↓	% unc.
Reference	-	-	8.86	3.00
PBMT-R	46.31	15.97	11.42	14.00
Hybrid	53.94	30.46	9.20	4.00
Transformer	49.85	27.92	8.00	15.03
DRESS	34.53	27.48	7.48	11.00
DRESS-Ls	36.32	27.24	7.55	13.00
NSELSTM-S	29.72	29.75	-	-
NSELSTM-B	53.42	17.47	-	-
EditNTS	23.87	32.35	5.47	0.00
SDISS(ours)	24.25	**34.06**	**4.58**	**0.00**

Newsela	BLEU	SARI ↑	FKGL ↓	% unc.
Reference	-	-	3.20	0.00
PBMT-R	18.19	15.77	7.59	5.85
Hybrid	14.46	30.00	4.01	**3.34**
Transformer	27.89	29.32	3.97	9.71
DRESS	23.21	27.37	4.11	11.98
DRESS-Ls	24.30	26.63	4.20	15.51
NSELSTM-S	22.62	29.58	-	-
NSELSTM-B	26.31	27.42	-	-
EditNTS	19.85	31.41	3.40	4.27
SDISS(ours)	18.81	**32.30**	**2.38**	5.01

Ablation Study

Newsela	**SARI ↑**	**FKGL ↓**	**%unc.**
SDISS	**32.30**	**2.38**	**5.01**
w/o copy loss	31.89	3.32	13.35
w/o LenTok	31.30	4.20	8.57
w/o GEncoder	30.24	3.56	9.07
w/o SenEncoder	31.55	3.98	10.32

Human Evaluation

	Newsela				WikiSmall				WikiLarge			
	F	A	S	avg.	F	A	S	avg.	F	A	S	avg.
Reference	4.29	2.84	3.73	3.62	4.28	3.78	3.24	3.77	4.26	4.09	2.62	3.65
PBMT-R	3.68	**3.64**	2.12	3.15	4.10	**4.04**	2.20	3.45	4.09	3.92	2.43	3.48
SBMT-SARI	-	-	-	-	-	-	-	-	4.03	3.72	2.52	3.42
DMASS+DCSS	-	-	-	-	-	-	-	-	4.12	3.75	2.76	3.54
Hybrid	2.98	2.62	2.74	2.78	3.41	3.64	2.50	3.18	2.93	2.66	**3.68**	3.09
DRESS	3.93	2.95	3.01	3.30	**4.29**	3.42	3.54	3.75	4.30	3.65	3.32	3.76
Transformer	3.81	2.72	3.04	3.19	4.25	3.62	2.43	3.43	4.32	3.97	1.89	3.39
EditNTS	**3.98**	3.02	3.23	3.41	4.09	3.12	**3.94**	3.72	4.40	4.03	2.98	3.80
SDISS(ours)	3.82	3.23	**3.32**	**3.46**	4.12	3.67	3.85	**3.88**	**4.62**	**4.13**	3.23	**3.99**

Case Study

Cases from Newsela	
Source	To prevent overfishing , the agreement would , among other things , make it much easier to establish marine protected areas -LRB- MPAs -RRB- in the high seas .
Reference	An agreement would make it easier to create marine protected areas .
PBMT-R	To prevent overfishing , the agreement would , among other things , make it much easier to establish marine protected areas -LRB- MPAs -RRB- in the high seas .
Hybrid	Overfishing to prevent the agreement would make it to establish areas .
DRESS	To prevent overfishing , the agreement would , among other things , make it much easier to establish marine protected areas .
EditNTS	The agreement would make it much easier to protected areas .
SDISS(ours)	The agreement would make it much easier to establish marine protected areas .

Cases from WikiSmall	
Source	A naval mine is a self-contained explosive device placed in water to destroy ships or submarines .
Reference	A naval mine is a bomb placed in water to destroy ships or submarines .
PBMT-R	A naval mine is a separate explosive device placed in the water to destroy ships and submarines .
Hybrid	A naval mine is a device explosive placed in water to destroy ships and submarines .
DRESS	A naval mine is a self-contained explosive device placed in water to destroy ships or submarines .
EditNTS	A naval mine is a explosive device can be .
SDISS (ours)	A naval mine is a explosive device placed in water to destroy ships or submarines .

PART 04

Conclusion

北京大學
PEKING UNIVERSITY

Conclusion

- We explore to incorporate semantic dependency graph into neural sentence simplification model. We propose a new model called SDISS, which can leverage the semantic dependency graph of input sentence to guide the simplification process.
- Our model generally achieves state-of-the-art performance on three benchmark datasets. Both automatic evaluation and human judgement indicate that our model improves semantic relevance.
- In the future, we will consider other semantic formalism like AMR and MRS to simplify sentences.

Thank you!