

# Generating Syntactically Controlled Paraphrases without Using Annotated Parallel Pairs

**Kuan-Hao Huang** and Kai-Wei Chang

University of California, Los Angeles

EACL 2021

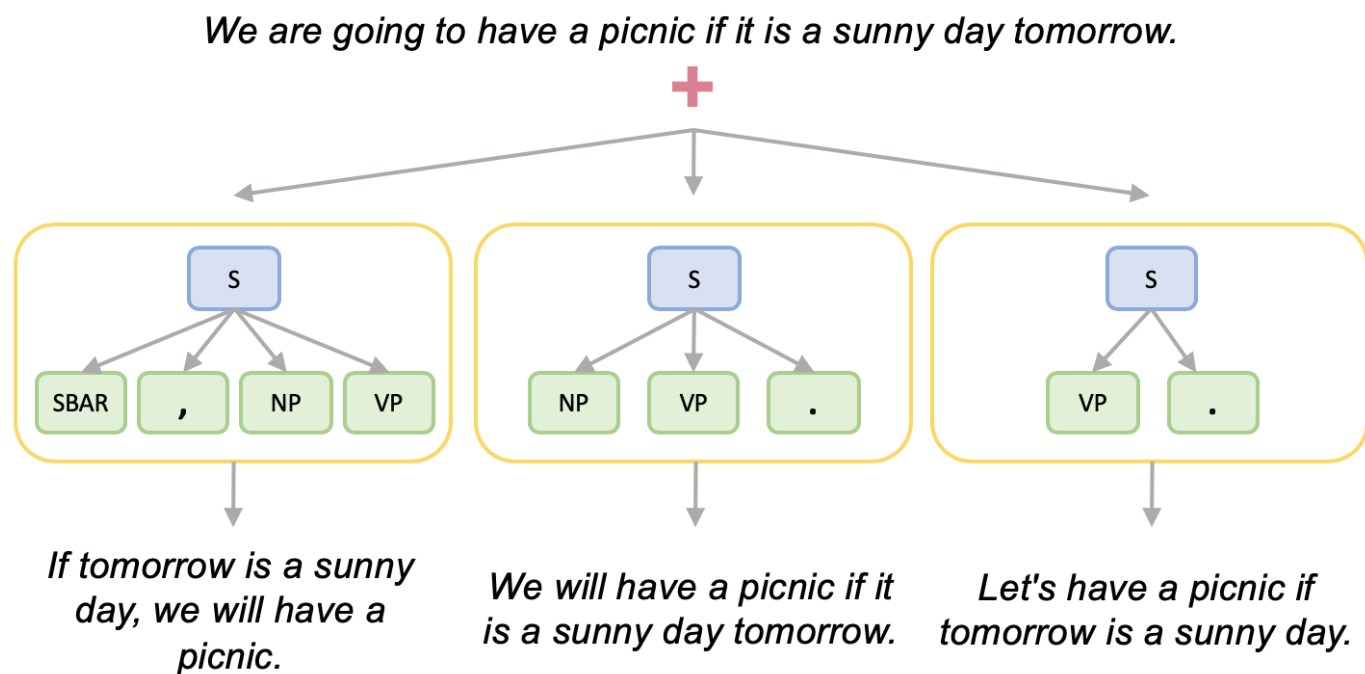


# Paraphrase Generation Tasks

- Build a paraphrase model
  - Input sentence → output paraphrase
- Supervised approaches
  - Need many **annotated paraphrase pairs** for training
- Unsupervised approaches
  - Generated paraphrases are **not diverse** in syntax
- Our goal
  - Generate **syntactically diverse** paraphrases
  - Train a paraphrase model **without using annotated paraphrase pairs**

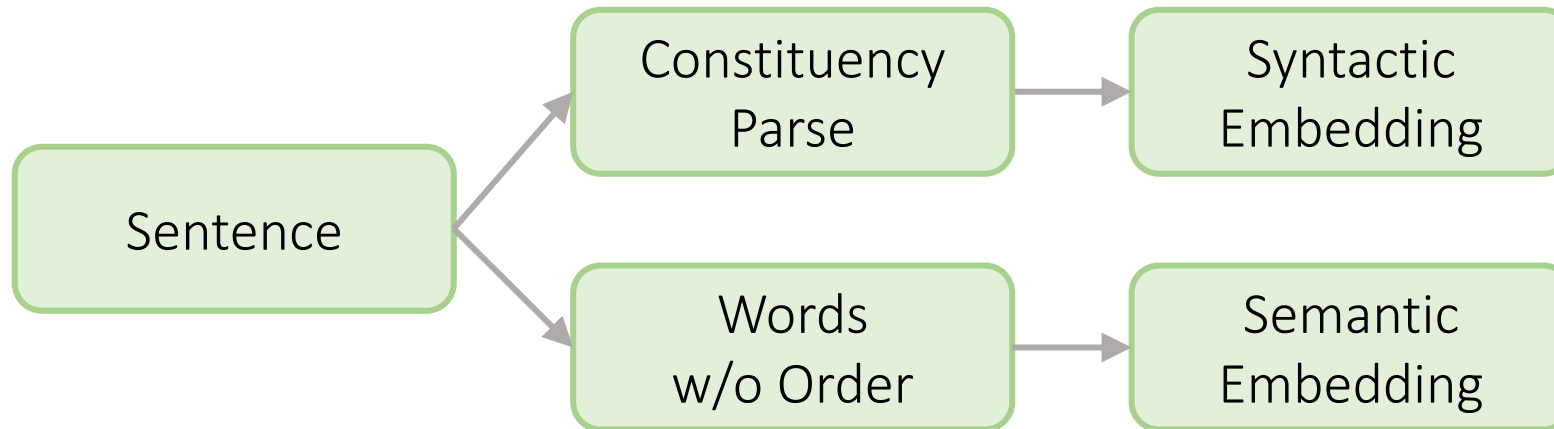
# Syntactically Controlled Paraphrase Generation

- Control the syntax of output paraphrases [Iyer+ 2018]
- Give model **target parse templates** as control signals
- Challenges
  - No ground truths
  - Syntactic control



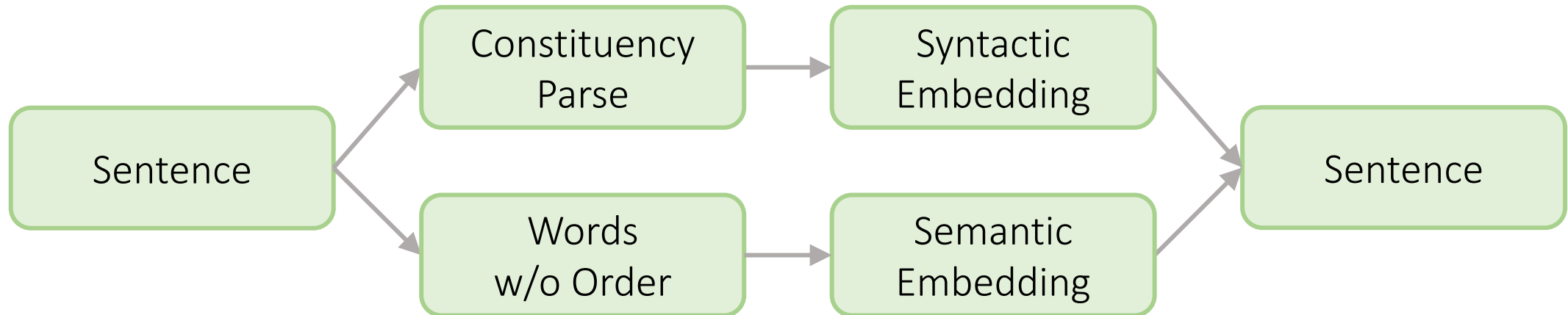
# Syntactically Controlled Paraphrase Generator (SynPG)

- Disentangle a sentence into syntactic and semantic embedding
  - **Syntactic embedding**: encoded from **constituency parse**
  - **Semantic embedding**: encoded from **bag of words**



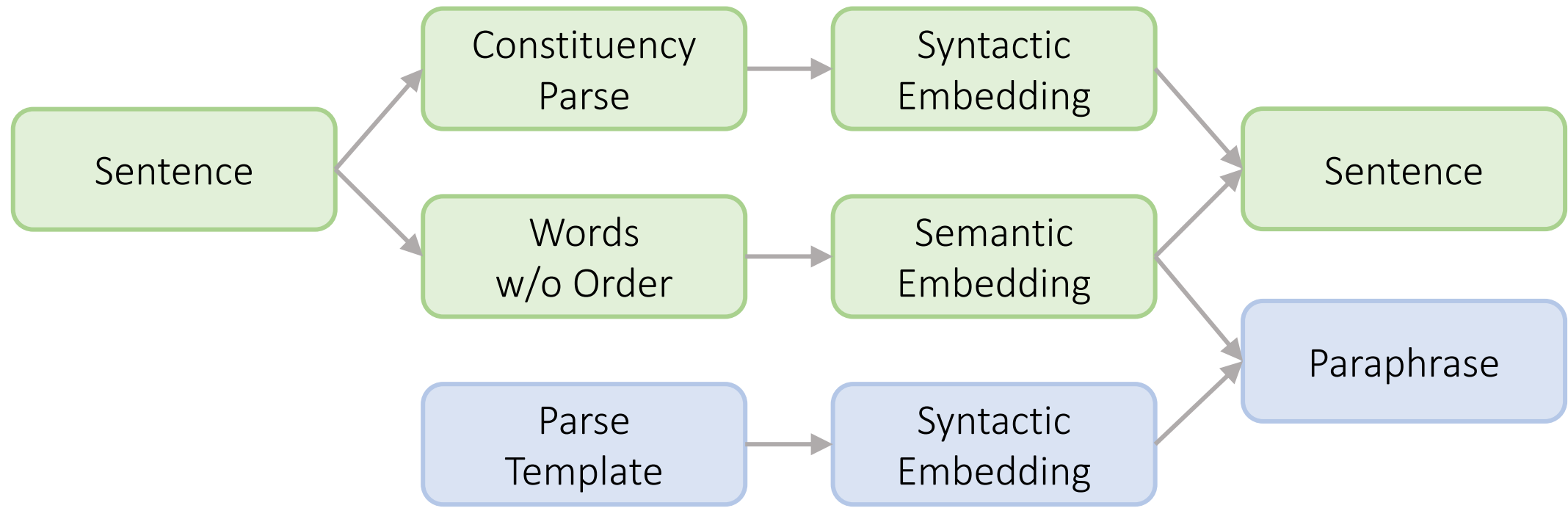
# Syntactically Controlled Paraphrase Generator (SynPG)

- Learn a decoder to reconstruct the input sentence
  - Reconstruction: no need for annotated paraphrase pairs
  - Disentanglement: ability to control syntax



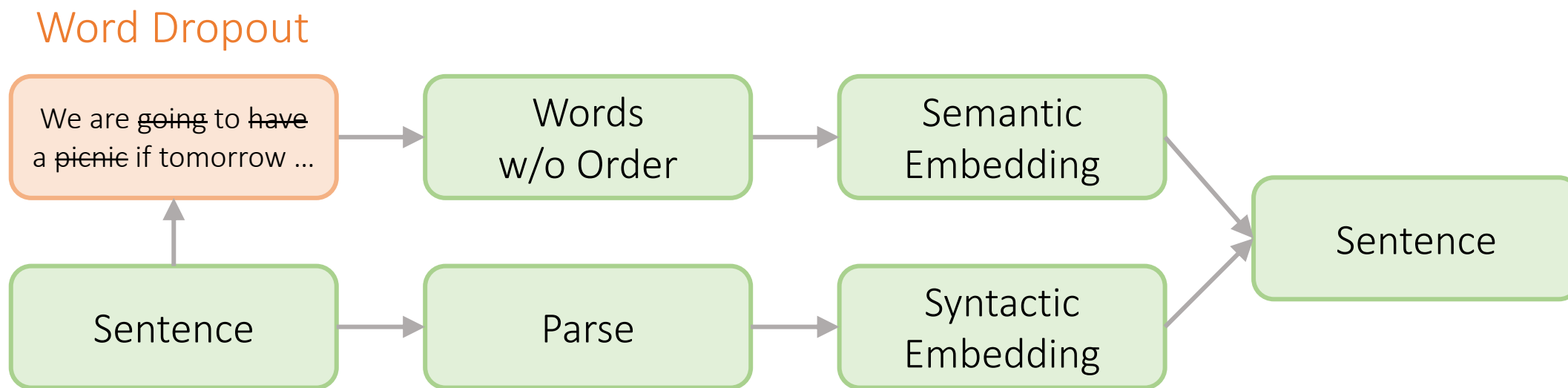
# Syntactically Controlled Paraphrase Generator (SynPG)

- Generate a paraphrase by replacing the syntactic embedding with the target parse information



# Word Dropout

- Add word dropout to encourage the model to generate words not appearing in the input sentence



# Evaluation on Syntactic Control Ability

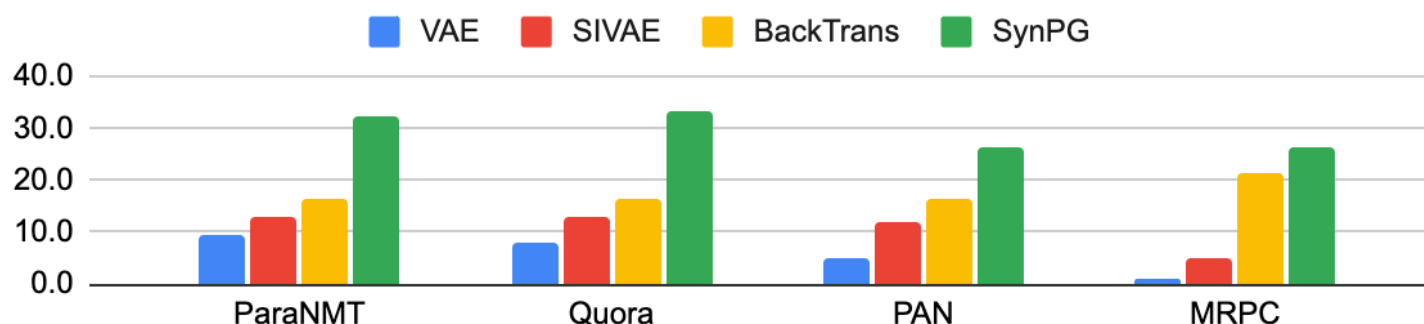
- Consider paraphrase pair  $(s_1, s_2)$ 
  - Get the parse  $(p_1, p_2)$
  - (input sentence, parse template) =  $(s_1, p_2)$
  - Ground truth is  $s_2$
- Evaluation metrics
  - **BLEU score**: similarity between the prediction and  $s_2$
  - **Template matching accuracy (TMA)**: how accurately the prediction follows  $p_2$  based on exact match on top-2 layers of parse tree
- Test on four paraphrase datasets (ParaNMT, Quora, PAN, and MRPC)



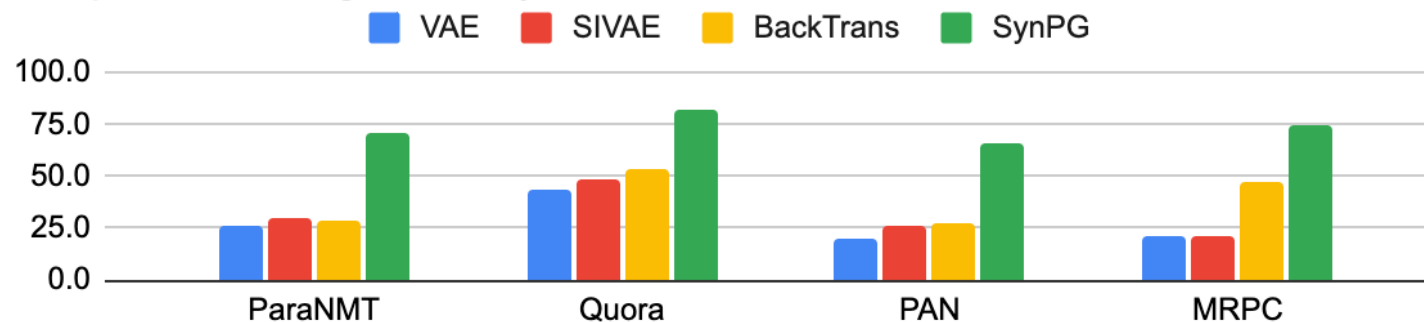
# Comparison with Unsupervised Models

- Compared models
  - Back Translation (BackTrans), Vanilla VAE (VAE), Syntax-Infused VAE (SIVAE) [Zhang+ 2019]
- Better syntactic control ability

BLEU score

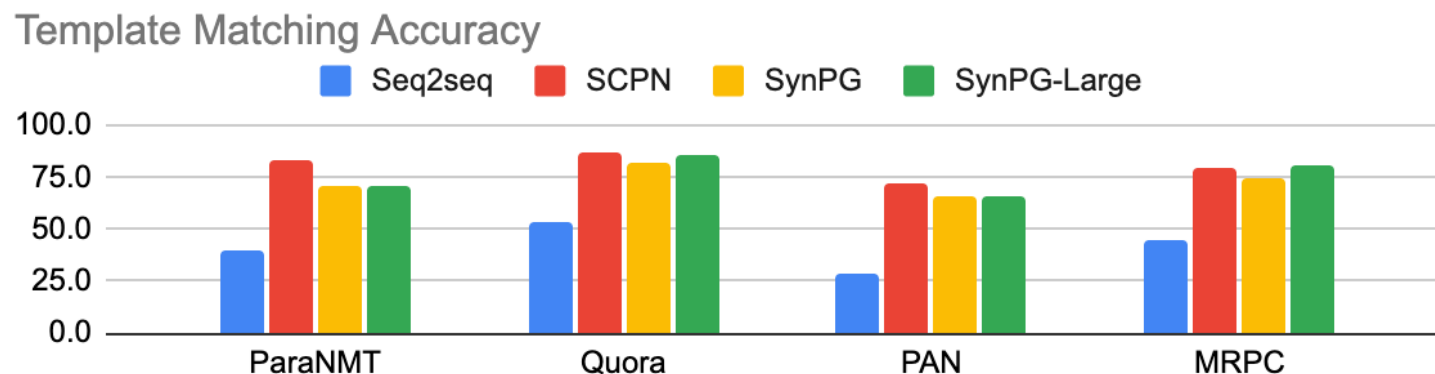
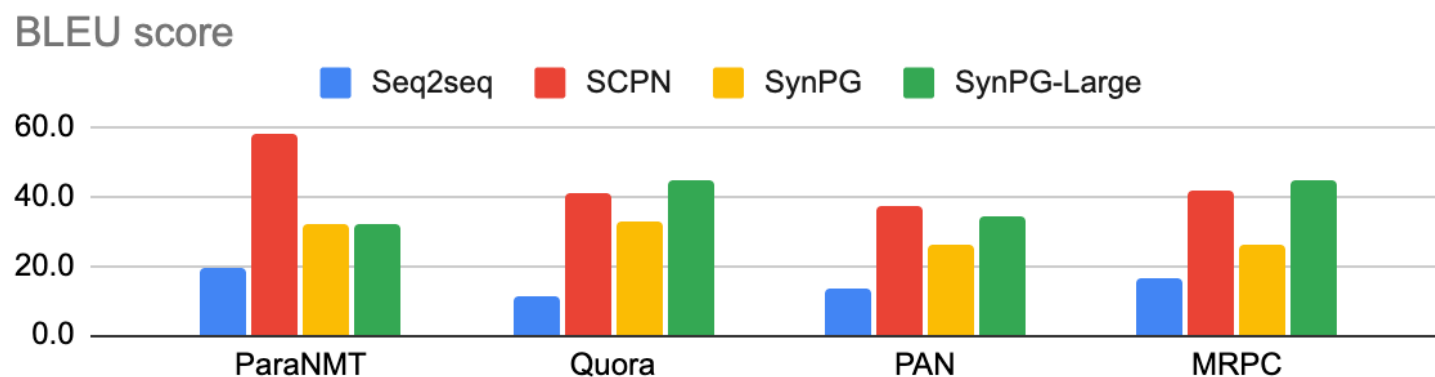


Template Matching Accuracy



# Comparison with Supervised Models

- Compared models
  - Seq2seq Model (Seq2seq), Syntactically Controlled Paraphrase Network (SCPN) [Iyyer+ 2018]
- **Competitive performance** to supervised models
- No need for annotated data → **boost performance by considering larger corpus**



# Paraphrase Examples

Template	Generated Paraphrase
<b>Original</b> (S (NP) (VP) (.)) (SBARQ (ADVP) (, ) (S) (, ) (SQ) (.)) (S (PP) (, ) (NP) (VP) (.))	can you adjust the cameras? you can adjust the cameras. well, adjust the cameras , can you? on the cameras, you can adjust them?
<b>Original</b> (SBARQ (WHADVP) (SQ) (.)) (S (``) (NP) (VP) (``) (NP) (VP) (.)) (S (ADVP) (NP) (VP) (.))	she doesn't keep pictures from her childhood. why doesn't she keep her pictures from childhood. " she doesn't keep pictures from her childhood " she said. perhaps she doesn't keep pictures from her childhood.

# Improve Model Robustness with Data Augmentation

- Focus on text classification tasks
- Generate syntactically paraphrases for data augmentation
- More robust against syntactically adversarial attacks [Iyyer+ 2018]

Model	SST-2		MRPC		RTE	
	Acc.	Brok.	Acc.	Brok.	Acc.	Brok.
Base	91.9	46.7	84.1	52.8	63.2	58.3
SynPG	88.9	<b>39.6</b>	80.1	<b>35.5</b>	60.7	<b>33.9</b>

# Conclusion

- We present SynPG to generate syntactically controlled paraphrases without using annotated paraphrase pairs
- SynPG uses a novel architecture to disentangle a sentence into semantics and syntax
- Extensive experimental results demonstrate the superiority of SynPG
- We show that SynPG can help to improve the model robustness



Code and pre-trained models are available at  
<https://github.com/uclanlp/synpg>

Thank You!