**To do before the presentation**

Discrete features (CRF)

1. SG2017 [train/dev/test]
2. WM2020 [train using SG2017, SG2017 + WM2020 (partial)]

BERT based fine-tuning

1. SG2017 – [train/dev (10~15% of train)/test]
2. WM2017 [train using SG2017, SG2017 + WM2020 (partial)]
3. Above with candidate sentence as either the first or the second sentence

Use IMHO and repeat 3. and 4.

Extract from BERT embedding + discrete features

1. Training via CRF
2. Training via LSTM? [we can discuss this]

For WM2020 data do the following

1. Use customized loss
2. Downsample a random 20% from the majority class

Discussion –

Why the performance is low for WM2020 (if it is…)

Claim vs. premise?

Domain difference?

Writing quality?

What are the ways to improve?

Joint learning? [show some results from SG2017]

Domain adaptation

Character embedding perhaps to take care of wrong spelling

DG: will provide more WM2020 data for training/test

DG: will prepare some more LM data

DG: will look at the joint learning for SG2017 data

DG: will provide the baseline (i.e., current WM) numbers