**TS**: pretrained encoder/decoder model

seq2seq

Let generate text, unlike BERT

Let "span denoising" objective

students opened their books

↓

<X> opened <Y>

**pretraining TS:**

Unmasked

XL

Unmasked

XL

SOS

<X> students <Y> their book

**fine-tuning TS:**
```
pretrained TS encoder

↑
this movie is good

pretrained TS decoder

↑
$<$sos$>$

softmax $d<\text{N}>

↑ positive
```