TS: pretrained seq2seq model
- unlike BERT, TS generates text
- encoder/decoder
- span corruption pretraining objective

Ex: "students opened their books"

$<X>$ opened $<Y>$
- fine-tuning:

  - parameter initialization needed
  - corrupted seq
  - this movie is good

  - missing spans

  - transformer encoder
  - transformer decoder

  - positive <Eos>

  - backprop

- the decoder is generating the text of the label, i.e. it needs to choose "positive" from the entire vocab.

- this is in contrast to BERT, in which you only produce \{positive, negative\} with a binary classifier