T5: pretrained seq2 seq model - Unlike BERT, TS generales text - encoder (decoder - span corruption pretraining objective ex: "students opened their books" 2 Sentitual totans < X> opened <Y> pretruing CX) Students CY) their books Transformer decode encoder <505> < x> Students < Y> <x> opened < y> their books barget is shifted over one token from input

Ane-turing.

Missing spans

Transforms

encoder

Correpted seq

Transform

encoder

Tr

- the decoder is generating the text of the label. i.e. it needs to choose "positive" from the entire vocab.
- this is in contast to BERT, in which you only produce Epositive, regative?

 with a binary classifier