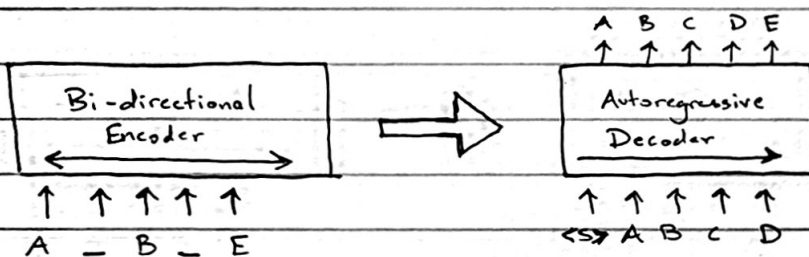


## BART

- Sequence to Sequence Transformer trained on this data: permute/delete/make tokens, then predict full sequence autoregressively



- \* BERT: Only parameters are an encoder, trained w/ Masked LM objective. Cannot generate text or do Seq2seq tasks

- \* BART: Both encoder and decoder. Can also use just the encoder wherever we would use BERT

## T5

- Pre-training: Similar denoising scheme to BART
- Input: texts with gaps.
- Output: A series of phrases to fill these gaps

Original text:

Thank you ~~for~~ ~~inviting~~ me to your ~~party~~ <sup>✓</sup> last ~~week~~.

Inputs:

Thank you <X> me to your party <Y> week.

Output (targets):

<X> for inviting <Y> last <Z>.