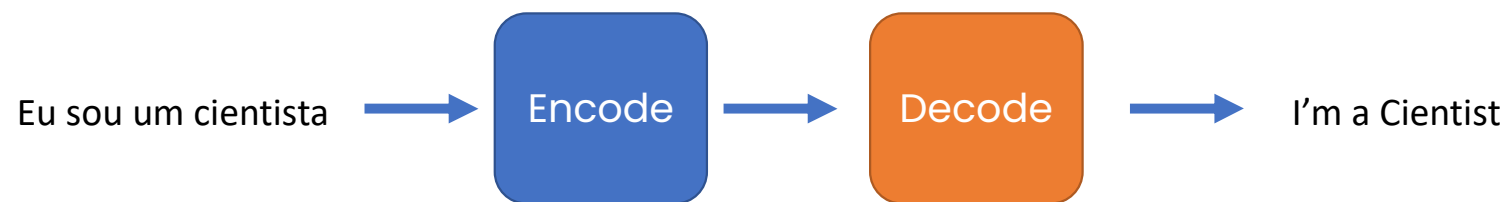


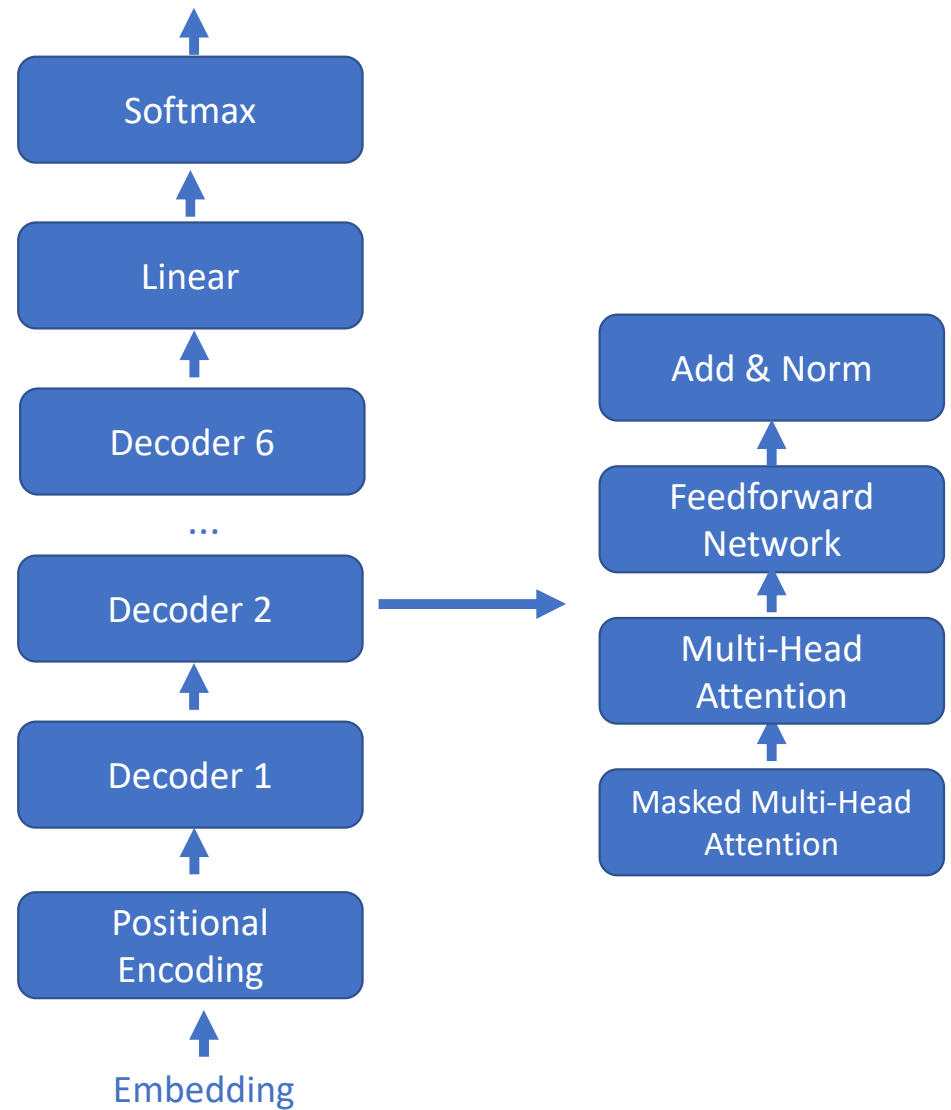
Decoder



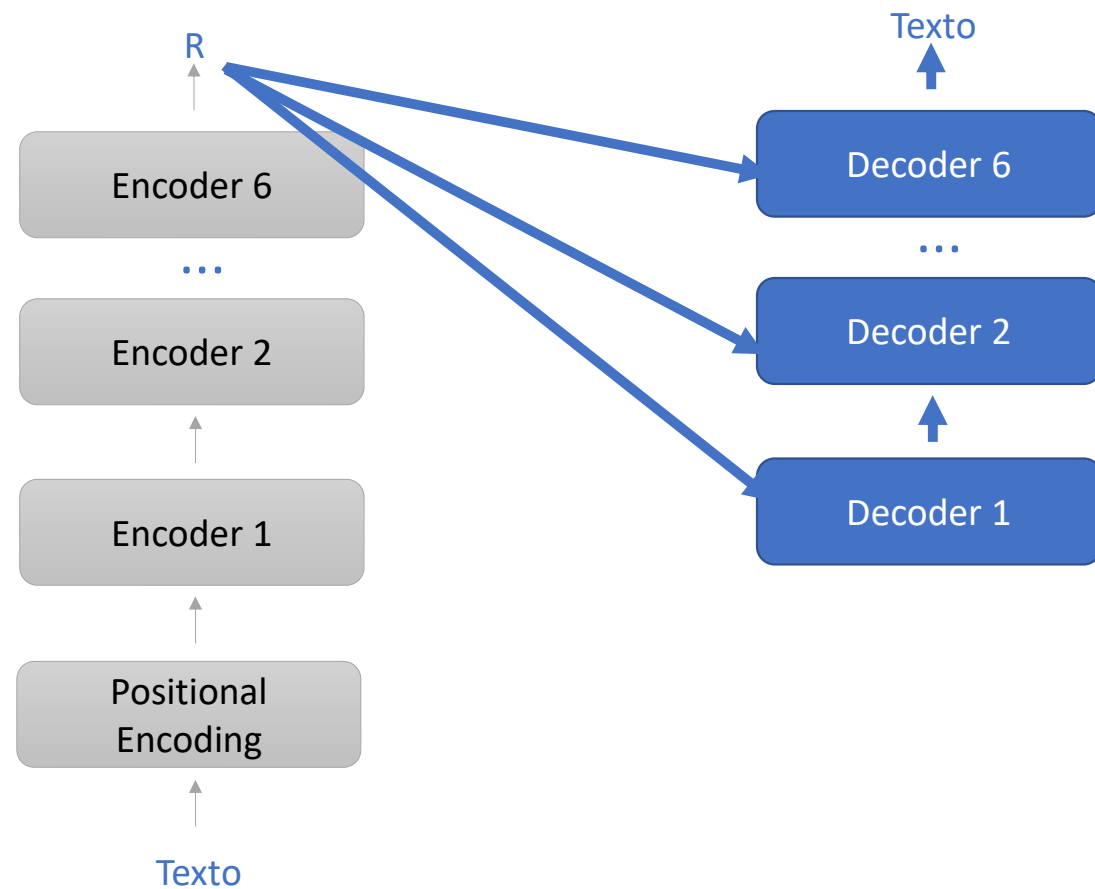
- Decoder

Também são camadas empilhadas

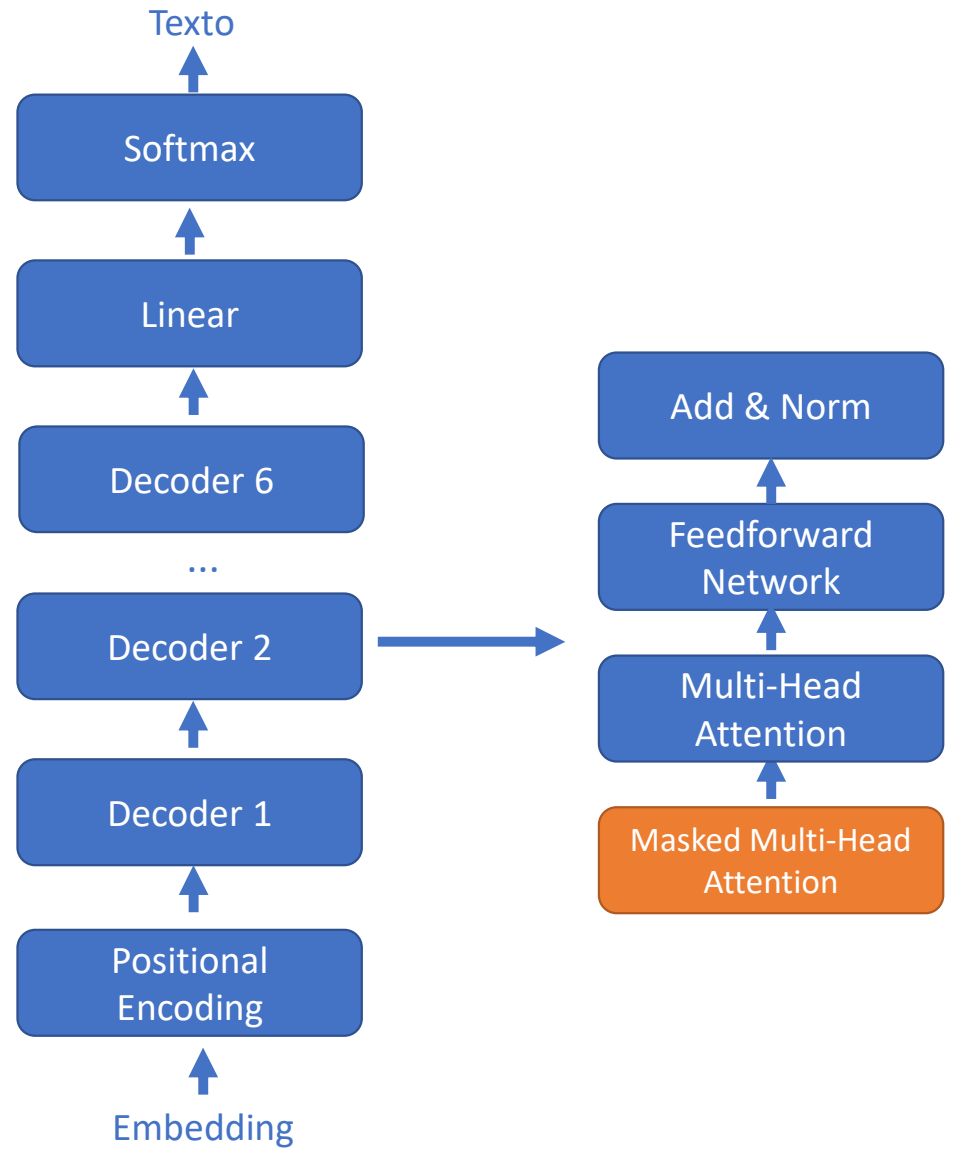
Também ocorre a conversão para embedding e positional encoding



- Cada Decoder recebe 2 entradas:
 - decoder anterior
 - a saída do encoder
- A conexão é com o elemento Multi-head attention do decoder



Decoder



Masked Multi-Head Attention

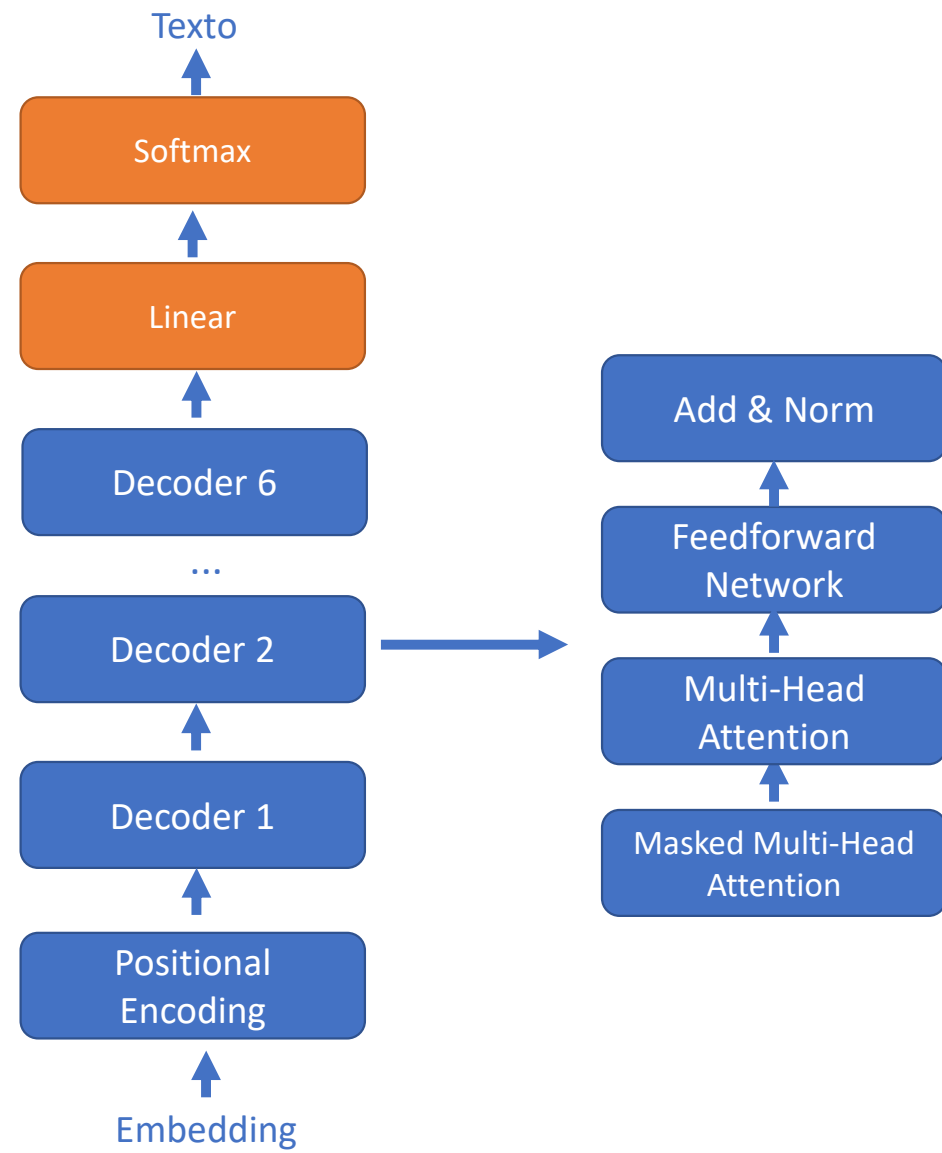
- `<sos>` Start of sentence
- `<eos>` End of sentence
- `<sos>` + “prever primeira sentença:” The
- `<sos>` + The “prever próxima:” animal
- `<sos>` + The + animal + ... + `<eos>`

Masked Multi-Head Attention

Palavras a direita são mascaradas, de forma incremental...

	The	Animal	Didn't	cross	the
The	0,9	---	---	---	---
Animal	0,57	0,79	---	---	---
Didn't	0,61	0,08	0,79	---	---
cross	0,79	0,39	0,60	0,39	---
the	0,60	0,57	0,39	0,57	0,39

Decoder



Linear e Softmax

- Linear produz um vetor com o tamanho do vocabulário
- Softmax produz as probabilidades
- Decoder gera as palavras com maior probabilidade