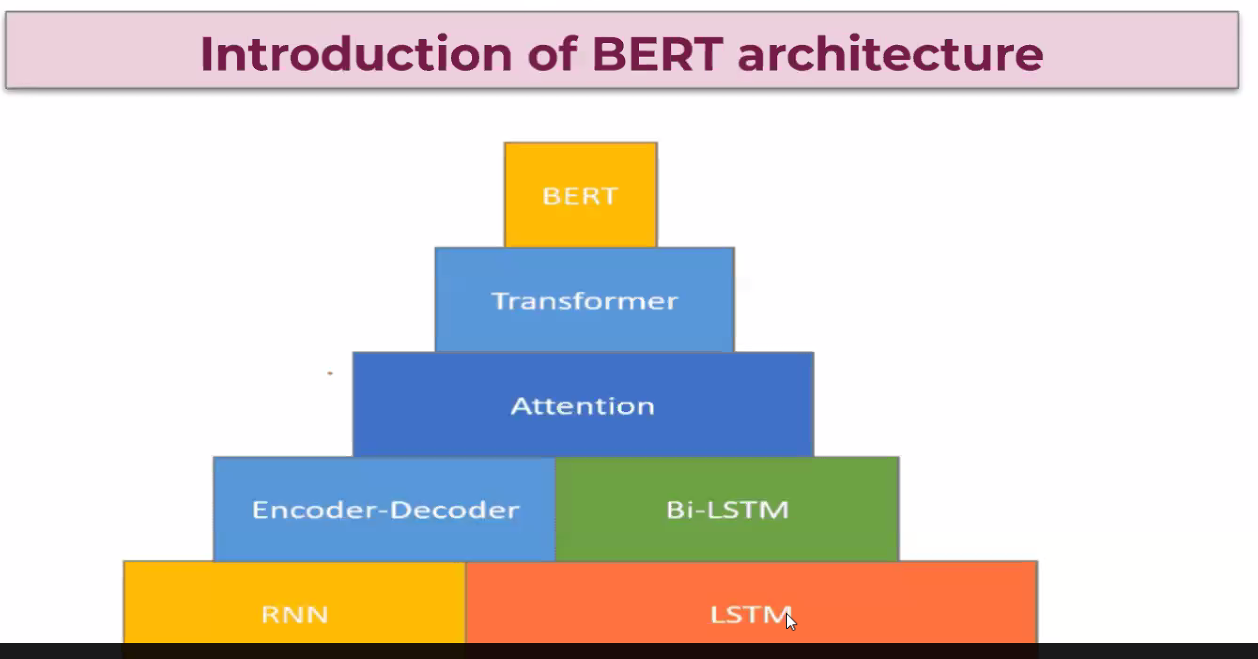
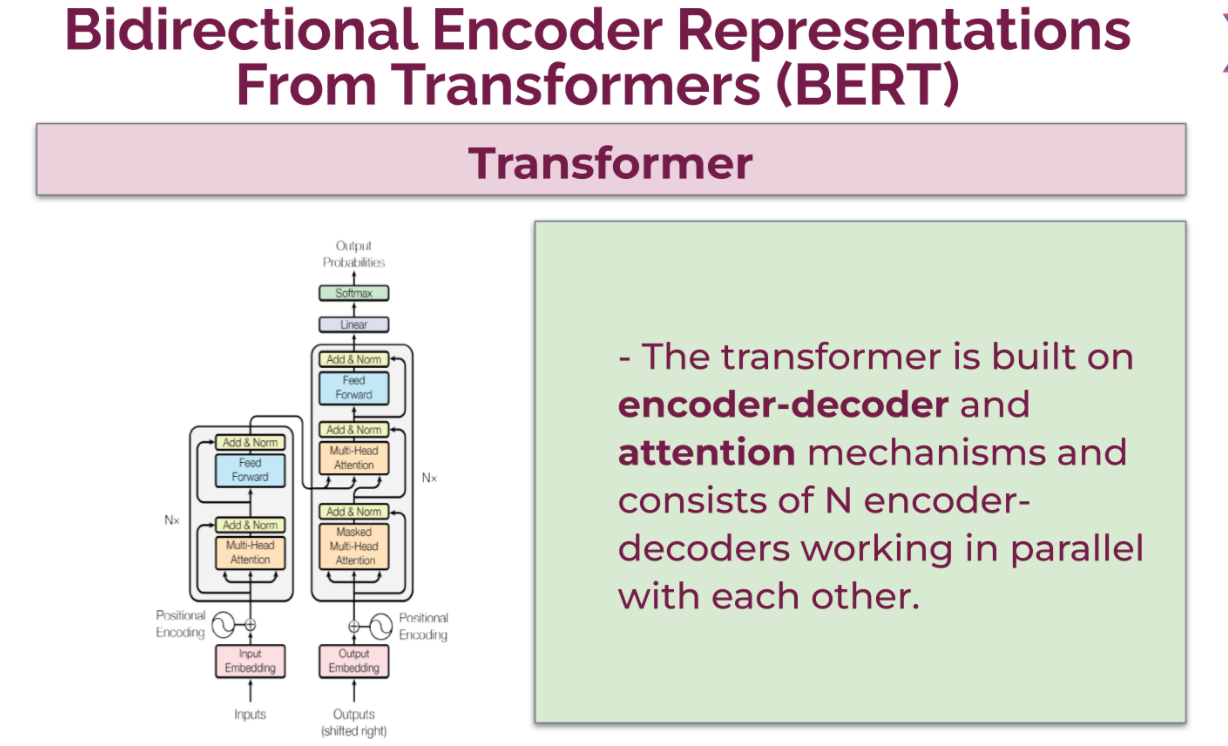
BERT



Encoder ile Türkçe bir dizin alıp Decoder ile onu İngilizce’ye çeviriyoruz.

Bi-LSTM çift taraflı çalışır.

Hem attention hem transformers kullanılarak da BERT modeli oluşturuluyor.



Sol taraf decoder, sağ taraflar encoder.

RNN modeller sırayla aldıkları için kelimeleri, sıralarını biliyor.

Mesela;

What time is it?

Başta what’I alacak sonra time sonra is gib

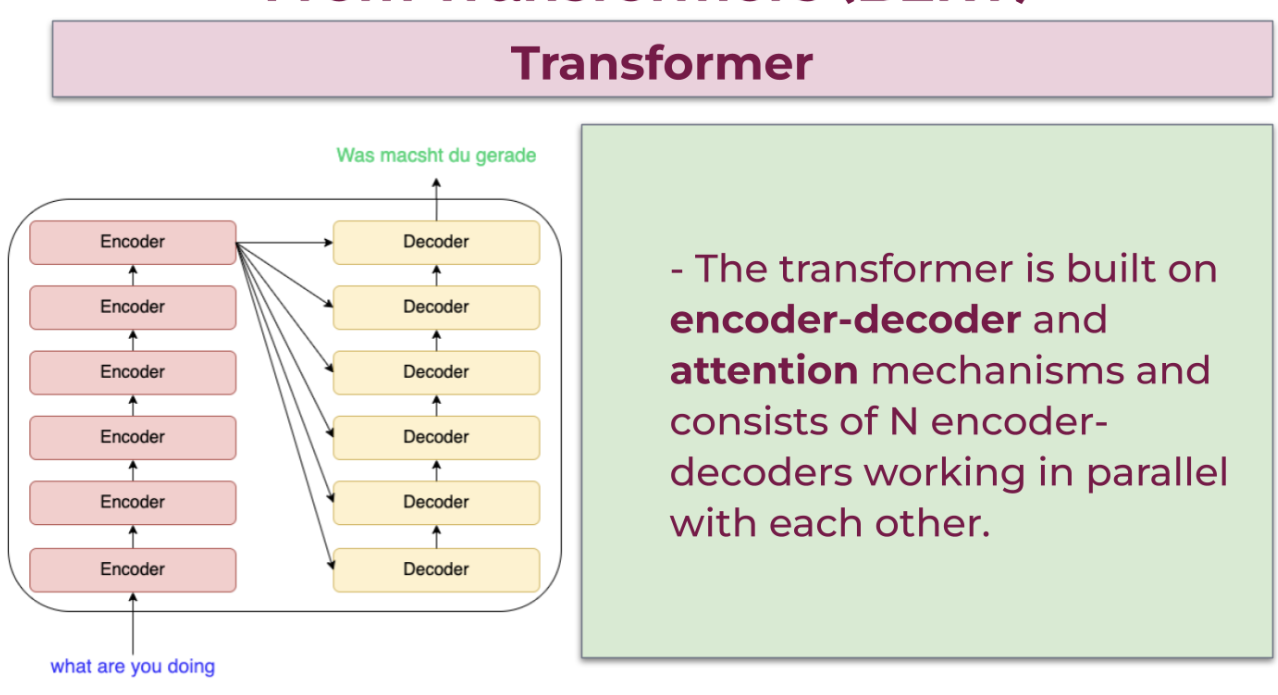
Lakin Transformers ise direct cümleyi bütün olarak alıyor. Position encoding işte bu sıralama işlemini yapıyor. Kelimelerin kaçıncı token olduğunu gösteren kod ekleniyor.

Bert eğitim aşamasında bazen 1’I kapatıp bazen 0’I kapatıp cevapları maskeliyor. Bu sayede tek girdiğimiz sınav üni sınavı değil, arada denemelere de giriyor modelimiz. Machine learning gibi değil yani. Biz machine learning’te hiç deneme yapmıyorduk eğitim sırasında. Bu işlemi yapan da 🡪 Masked Multi-Head Attention

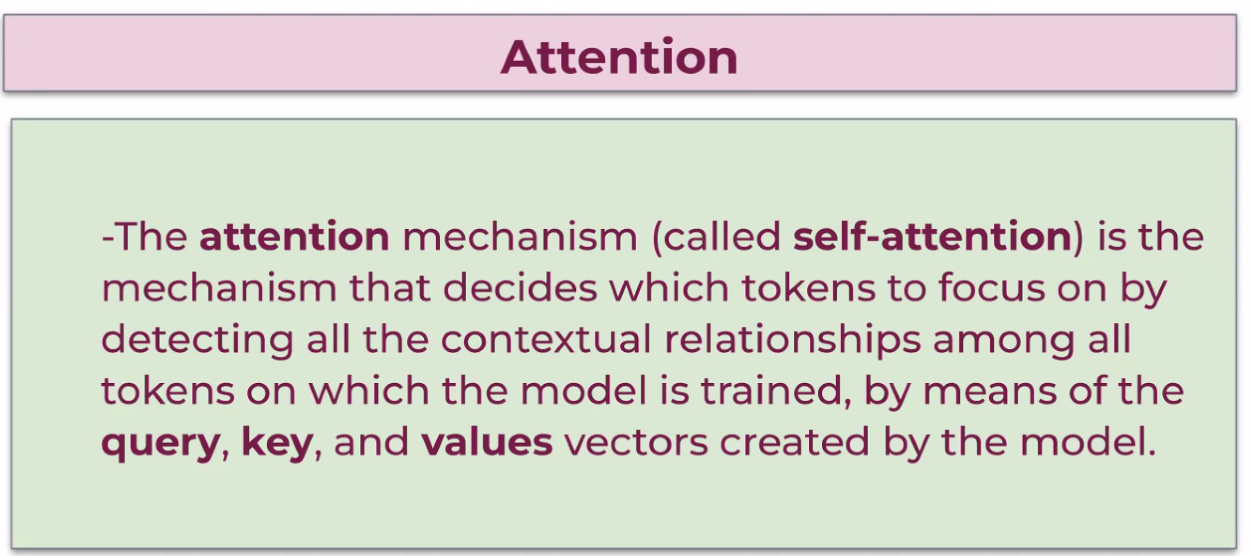
Bert modeler 512 token’e kadar sınırlı.

Şuan için piyasadaki problemleri halletmek için 512 token gayet enough.

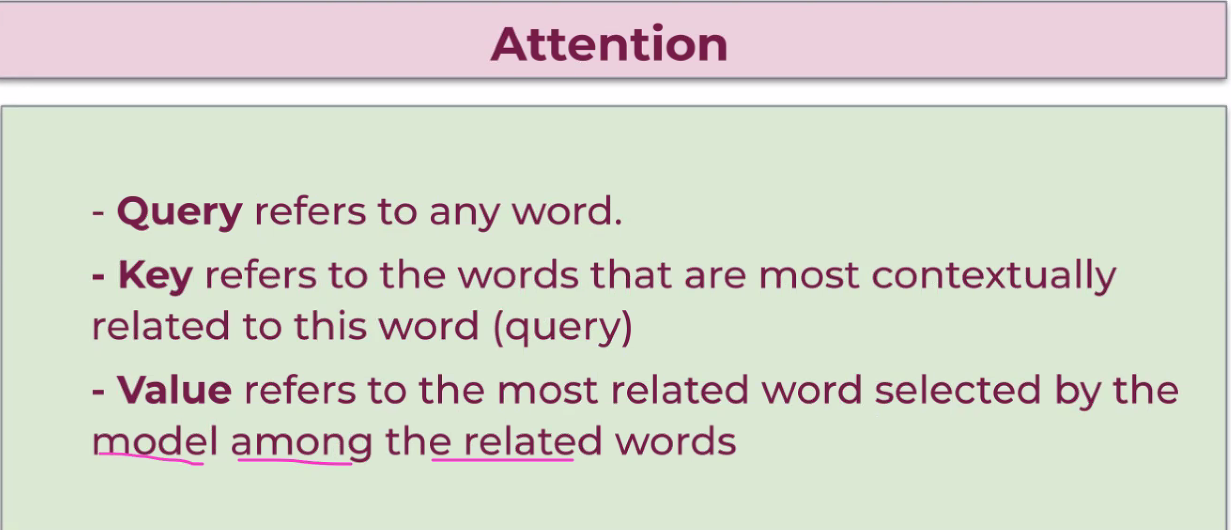
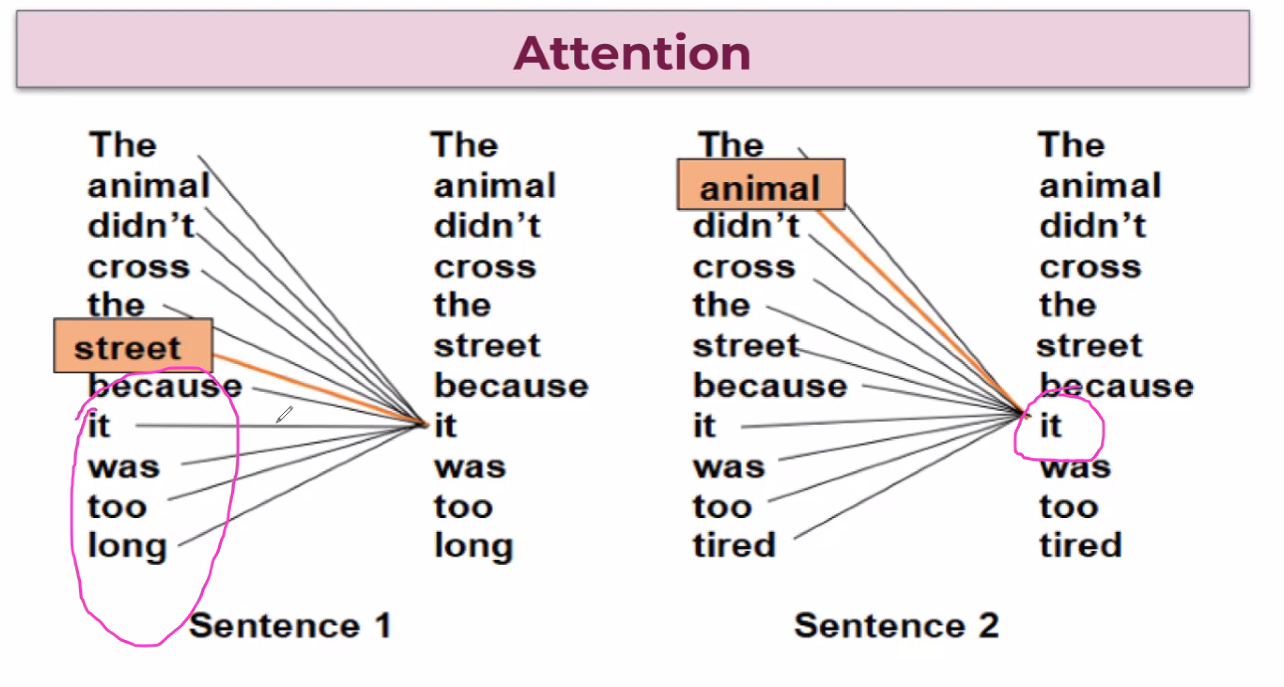
Bert zaten hazır bir model, biz ufak ayarlamalar yapacağız sadece.



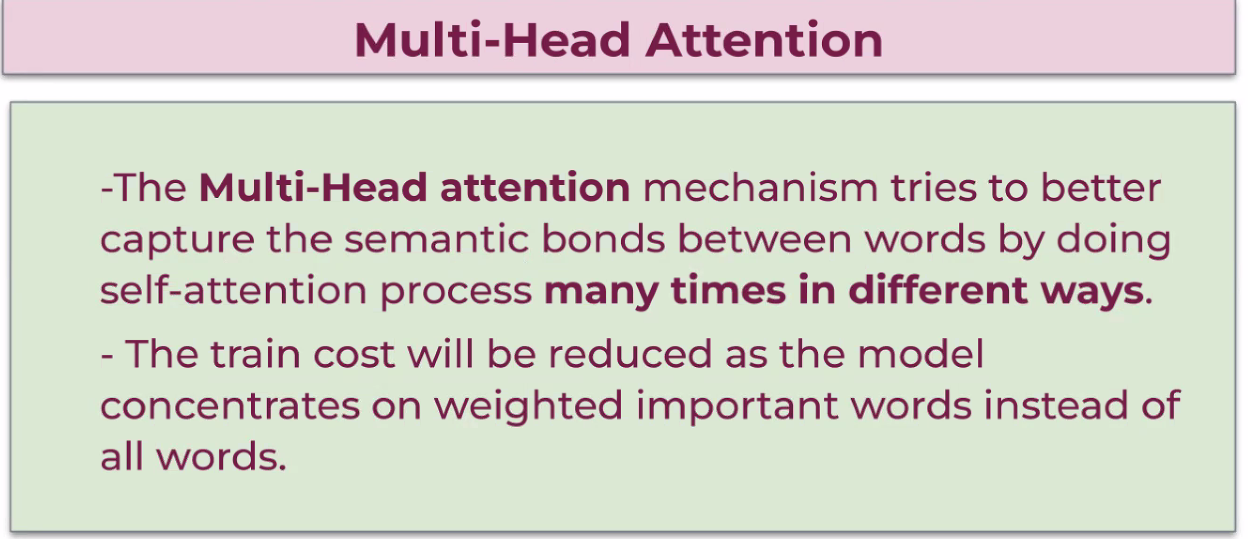
Bu şekilde hem 6 tane encoder, 6 tane decoder’ımız var.

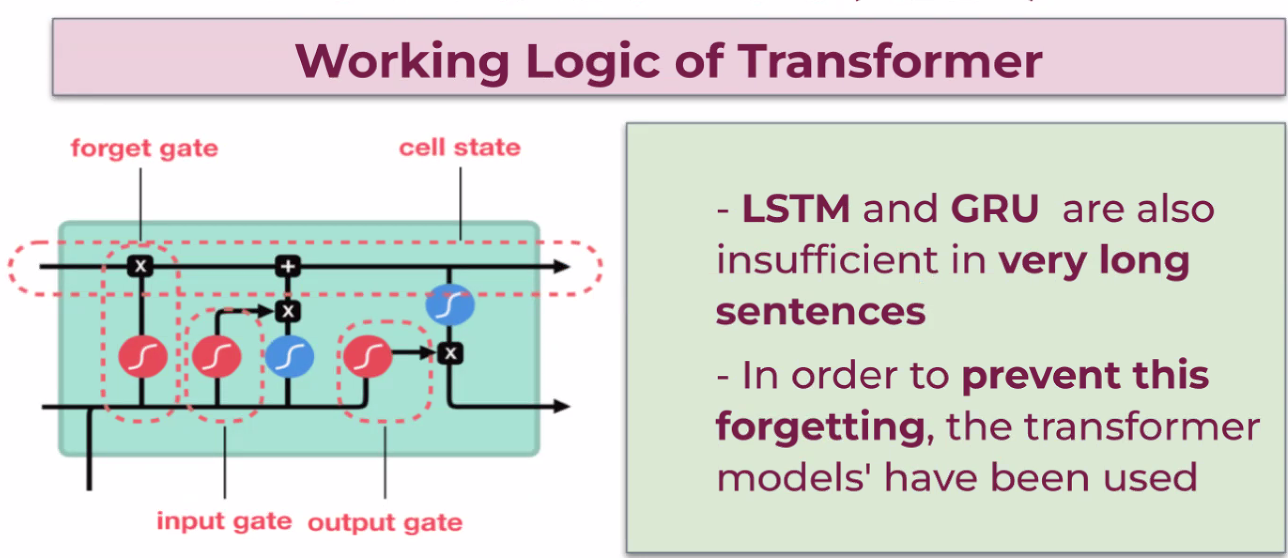


Attention bir nevi cümlenin ögelerine ayırıyor, sorular soruyor ve anlamsal ilişkileri bu sayede çok iyi buluyor. Multi head attention dediğimiz işte bu.

Attention 3 vektörümüz var: Query, key, and values.

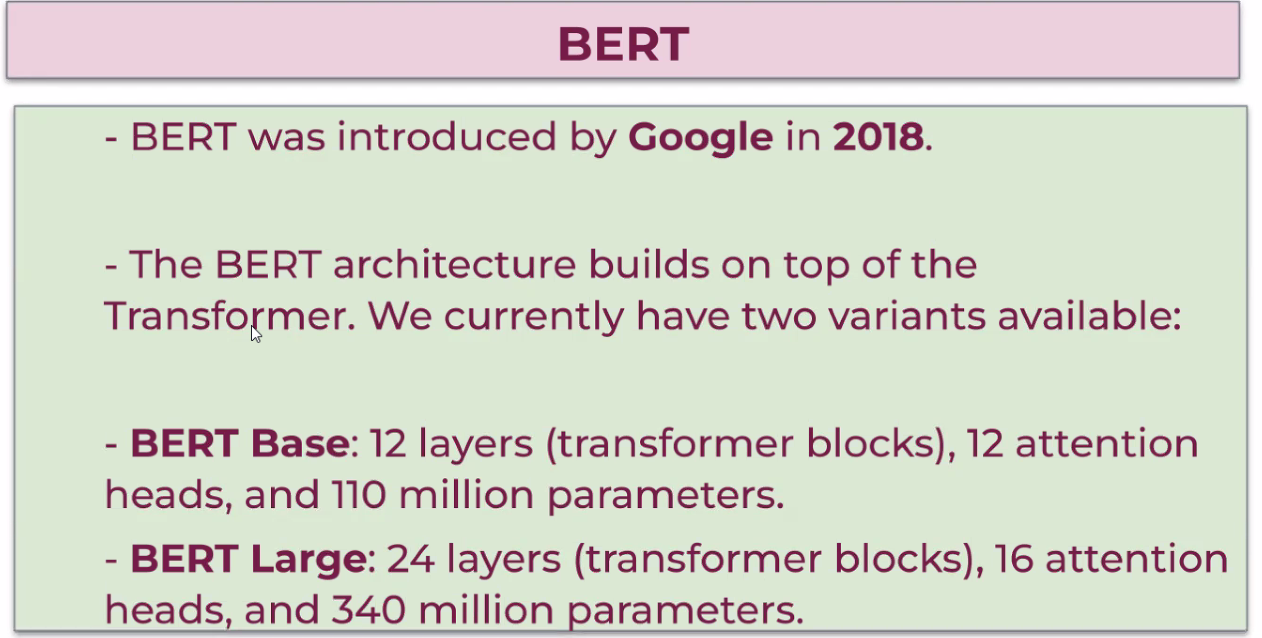
Mesela bir nesnenin yorgun olamayacağını biliyor.



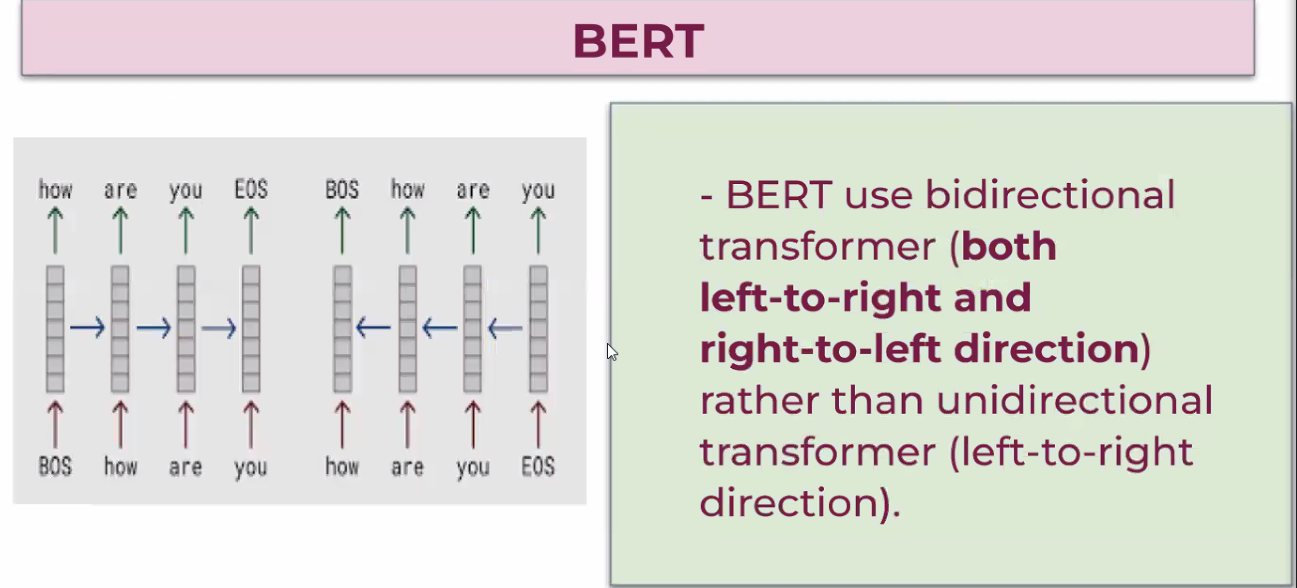


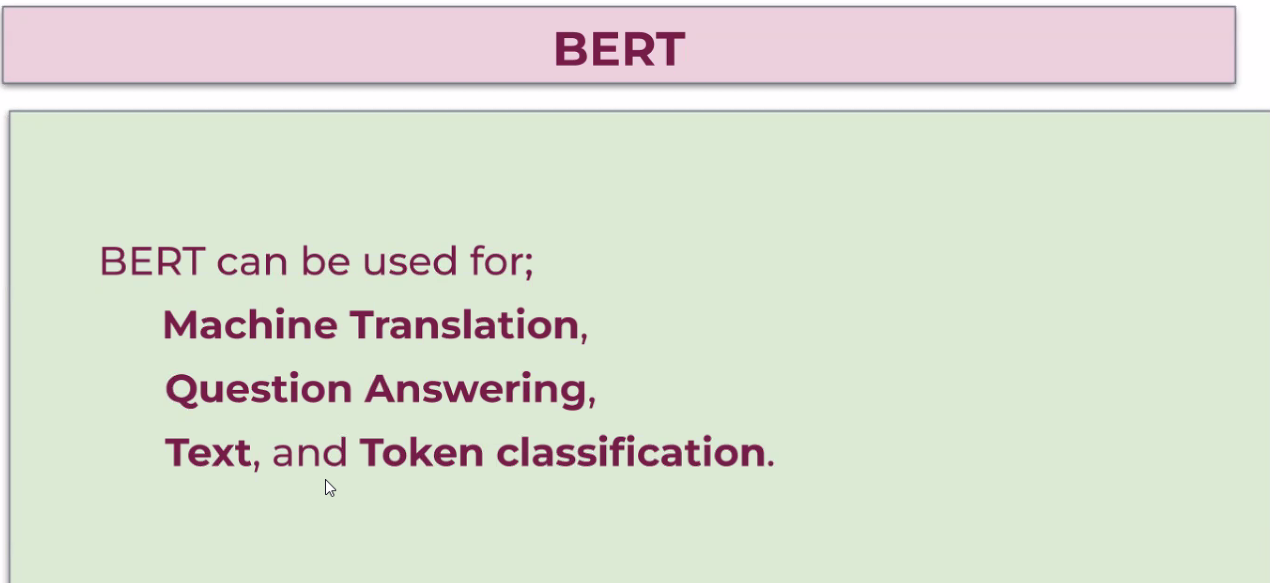
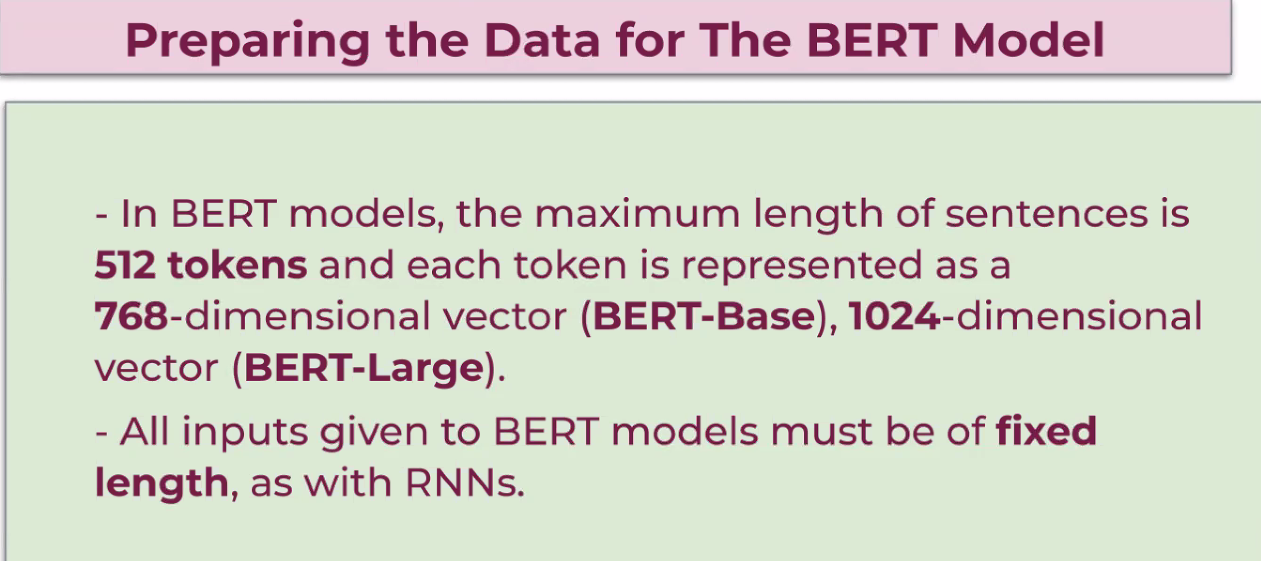
LSTM ve GRU’nın yetersiz kaldığı yerlerde Transformerslar devreye giriyor.

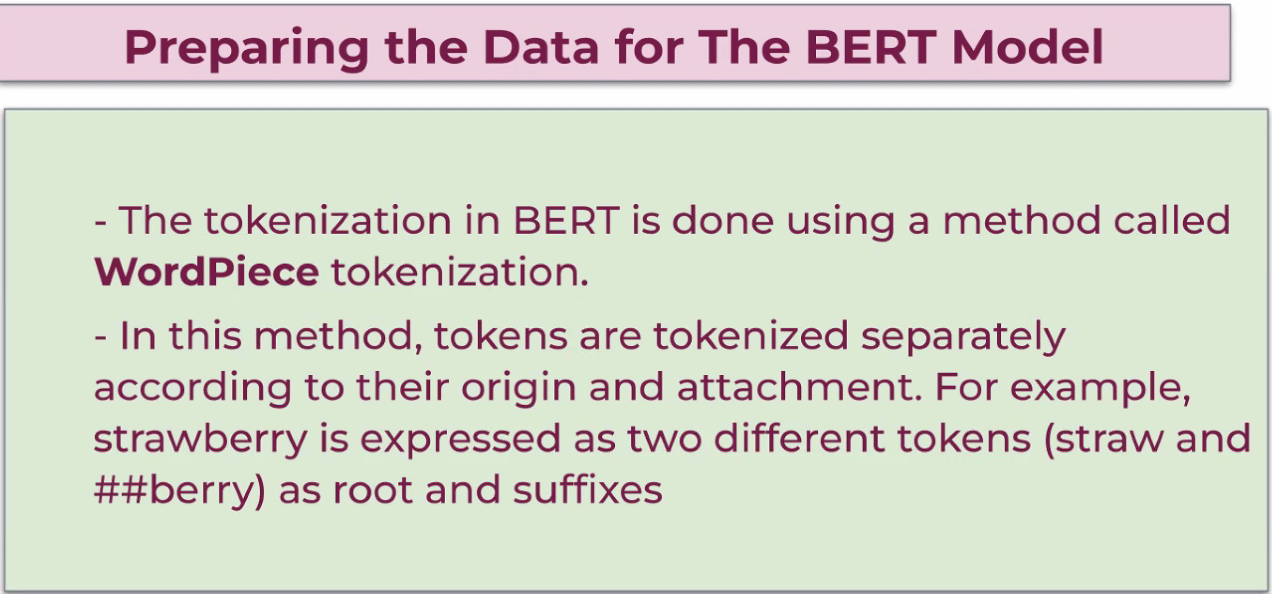
Bu kadar iyi çalışmasnın sebeplerinden beri de, mesela RNN’de biz dense layerda sadece son hidden layerı aktarıyorduk. Ama encoder kendisindeki tüm hidden layerları decoder’a gönderiyor. Bu yüzden de uzun süreli hafızamız oluyor aslında.

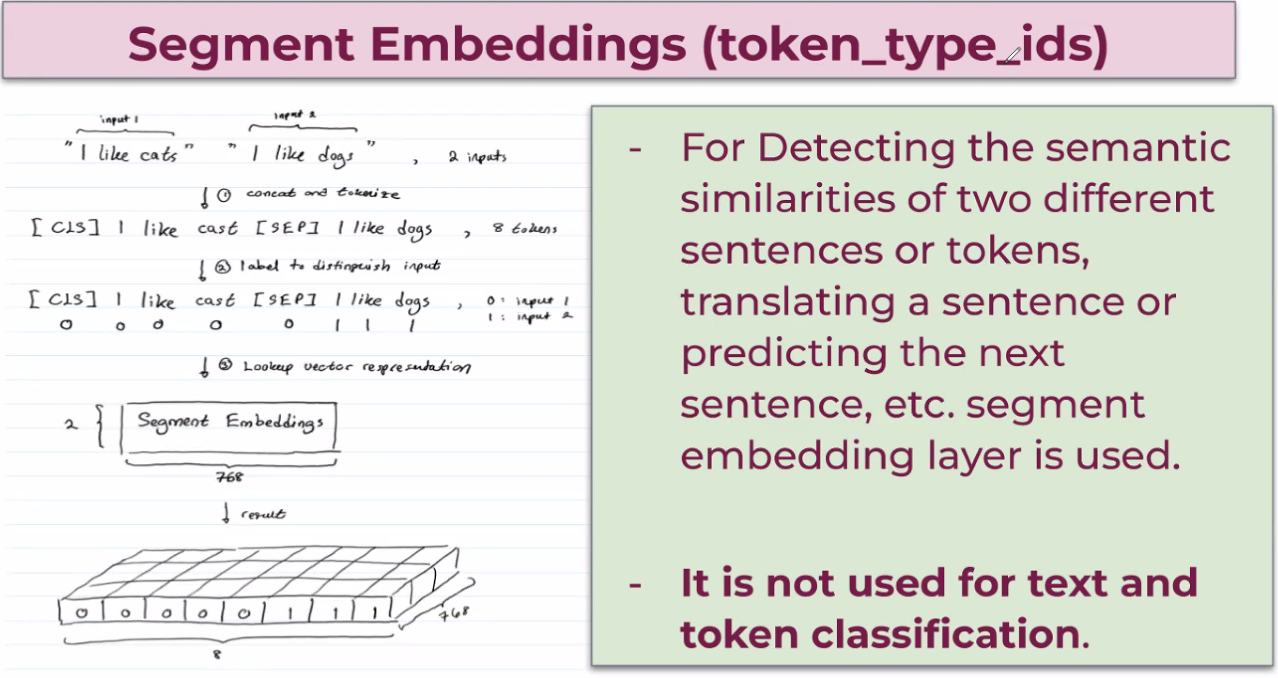


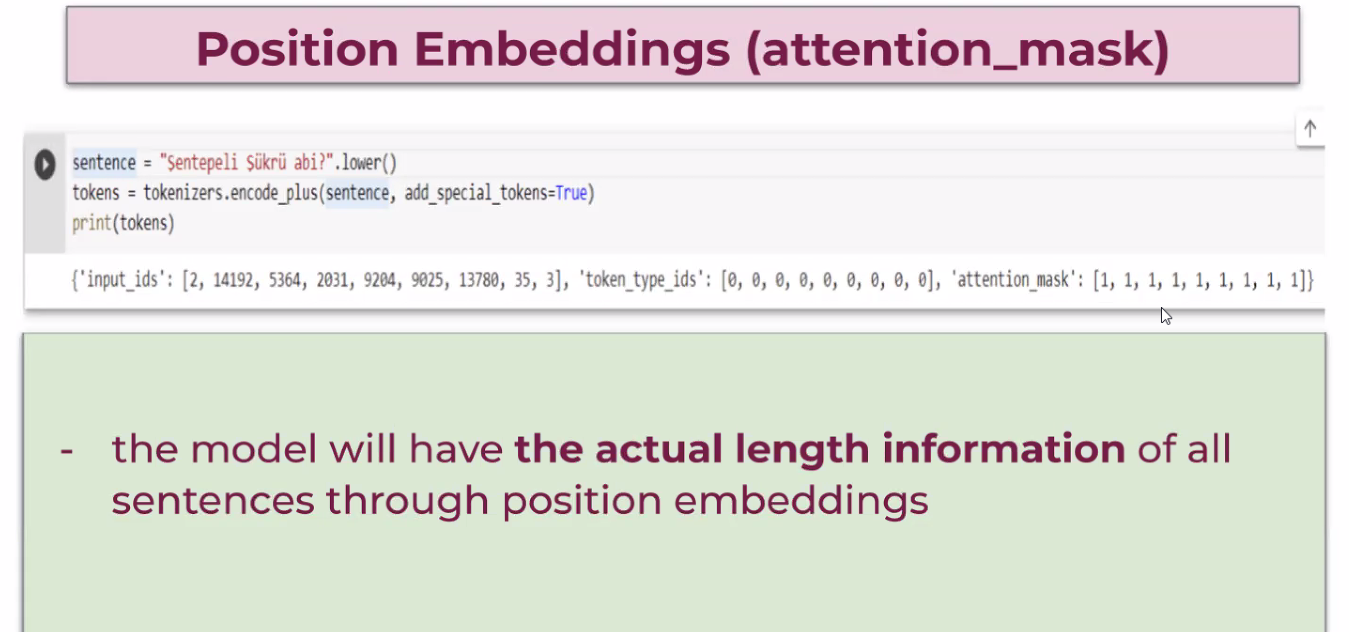












 What is the tokenization method used in BERT models? What is this method used for?

A: In BERT models, tokenization is done with the wordpiece method. In this method, tokens are tokenized separately according to their origin and attachment. For example, strawberry is expressed as two different tokens (straw and ##berry) as root and suffixes. let's look at another example. Here, embeddings is expressed as em, ##bed, ##ding and ##s. In this way, words/tokens that are not in the model's vocabulary (size of BERT vocabulary:30.522)  are also represented by a vector.

Q: How should the input be given to the model in BERT models?

A: n BERT models; The input must be given to the model with a tensor consisting of token, segment, and position embeddings vectors. Otherwise, the model will not work.