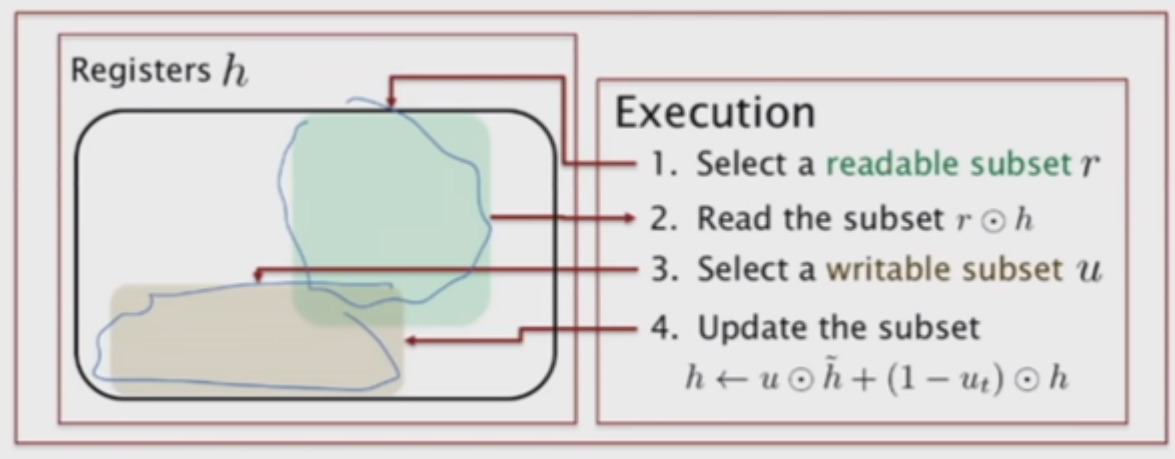
Gru and LSTM

GRUs and LSTMs work on the basis that we selectively choose a certain part of the hidden state which we update and certain parts that we don't. Visual Representation (of GRU):



The reason why we don't get vanishing and exploding gradients in GRUs and LSTMs is because we selective divide the hidden state into parts that we want to update, parts we want to throw away and parts we want to keep and after doing the operations to the matrices we add the individual ones which have the elements we want to retain, ones we wanted to add and once we didn't want. This adding together is the reason for elimination of vanishing and exploding gradients.

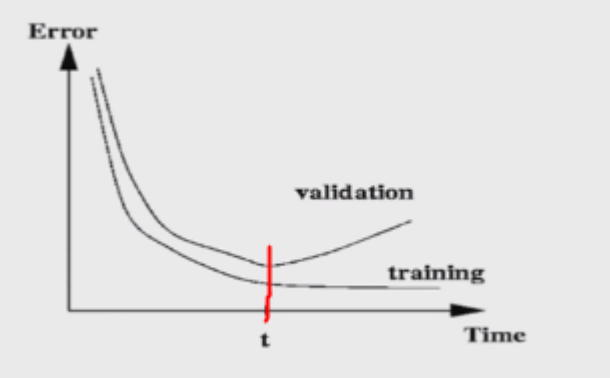
Remember: UNK - "Unidentified token" is used to replace words that are out of the model's vocabulary.

MT evaluation:

Even today, the best possible method of translation is through a human evaluation, especially giving humans two outputs and asking them which is a better translation. BLUE when it came out was very god but once people started making models that specifically aimed at getting better blue scores then its validity dropped massively.

Preventing overfitting:

We can prevent overfitting by continuously training a model to a particular training data but having a validation dataset separately with which we validate the output and once the performance on validation data decreases, that is where we stop our training.



Stop at time t. Although overfitting is not always bad, we sometimes have to find and remember specific cases to be good. **Good models overfit on their training data, almost to a level of getting 100% accuracy on the training data.**

Loss and Learning rate: