Applied Text Analytics & Natural Language Processing

with Dr. Mahdi Roozbahani
& Wafa Louhichi

Deep Learning
Long Short-Term Memory (LSTM) + Attention

Some of the slides are based on Ming Li (University of Waterloo – Deep Learning Part)
with some modifications

Learning Objectives

In this lesson, you will learn how to augment LSTM with an attention structure LSTM

- Why we need Attention Mechanism
- Encoder-Decoder architecture
- How it works

Language Translation Example

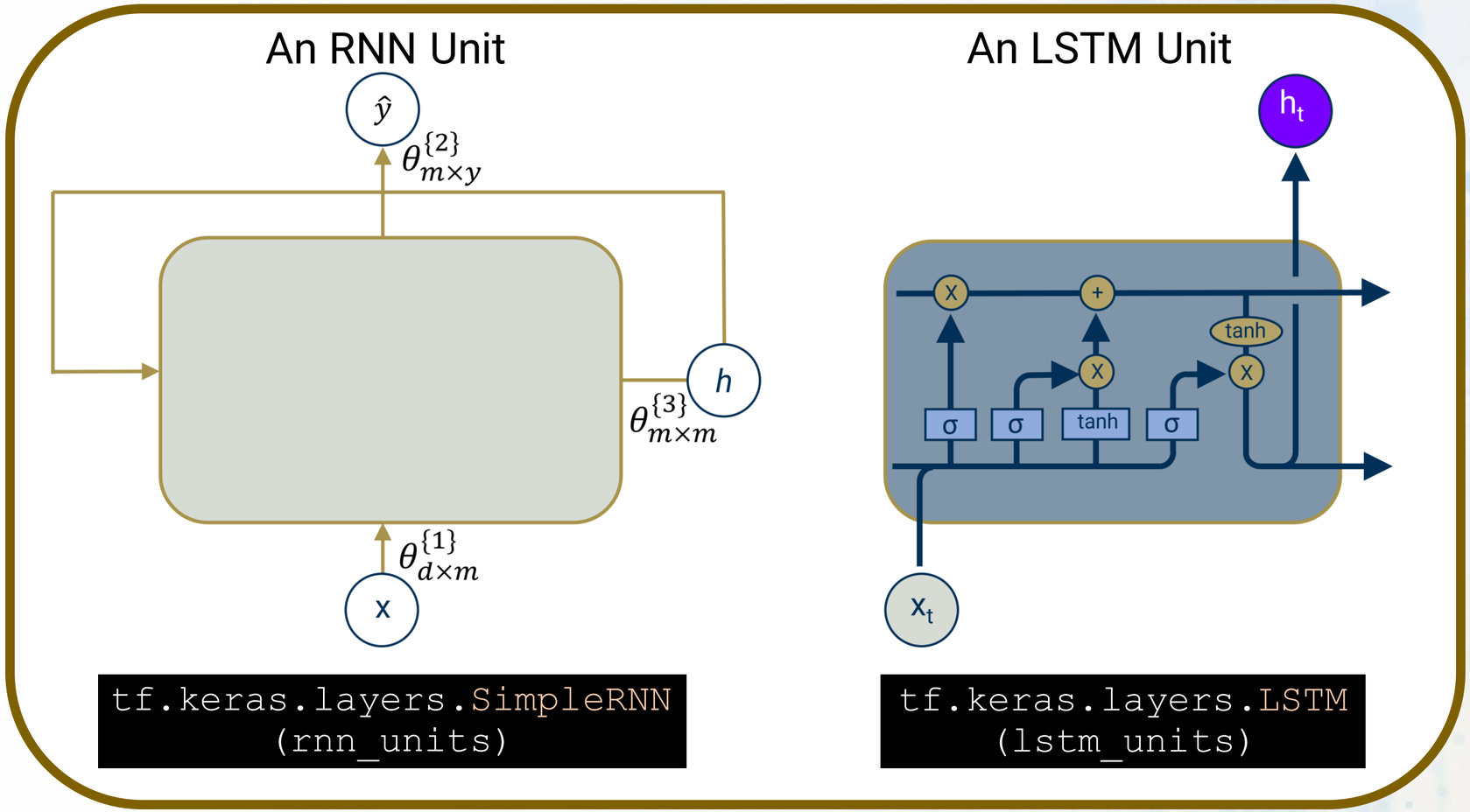
Let's say we are going to translate "How are you" from English into Polish:

English: *How are you*

Polish: *jak się masz*

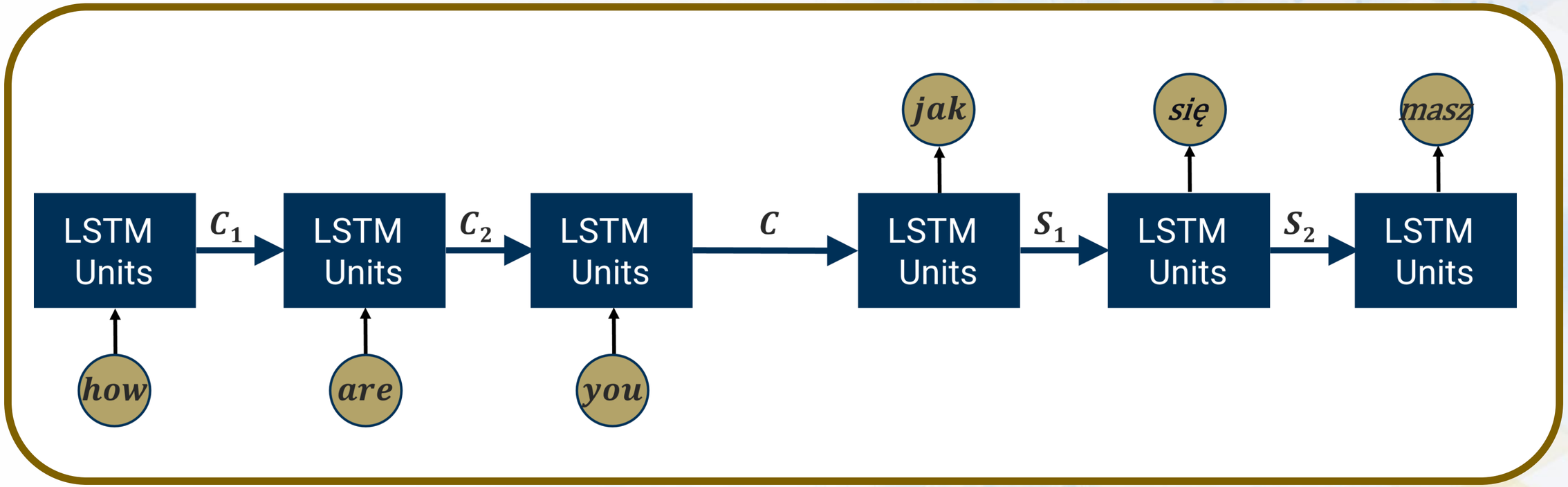
We are going to use LSTM units to perform this task.

RNN and LSTM Units

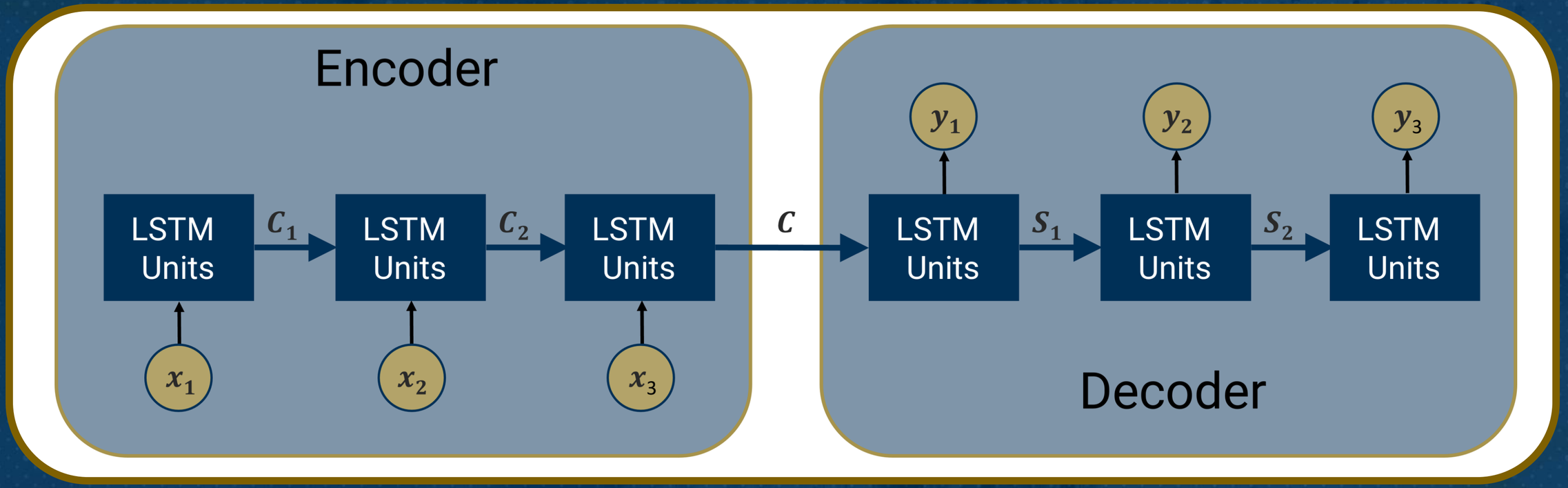


LSTM essentially augments the RNN unit by creating gates that allow some information to be passed on through the network and some to be forgotten.

Back to our Language Translation Model



Encoder and Decoder

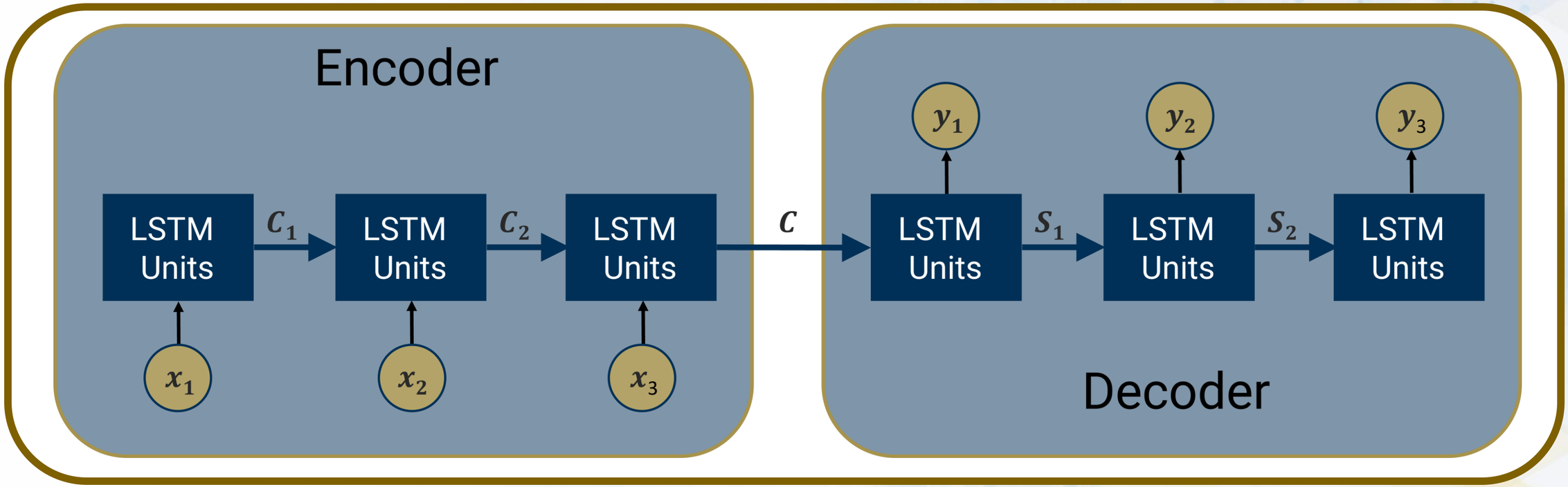


C_i: It is the encoder state

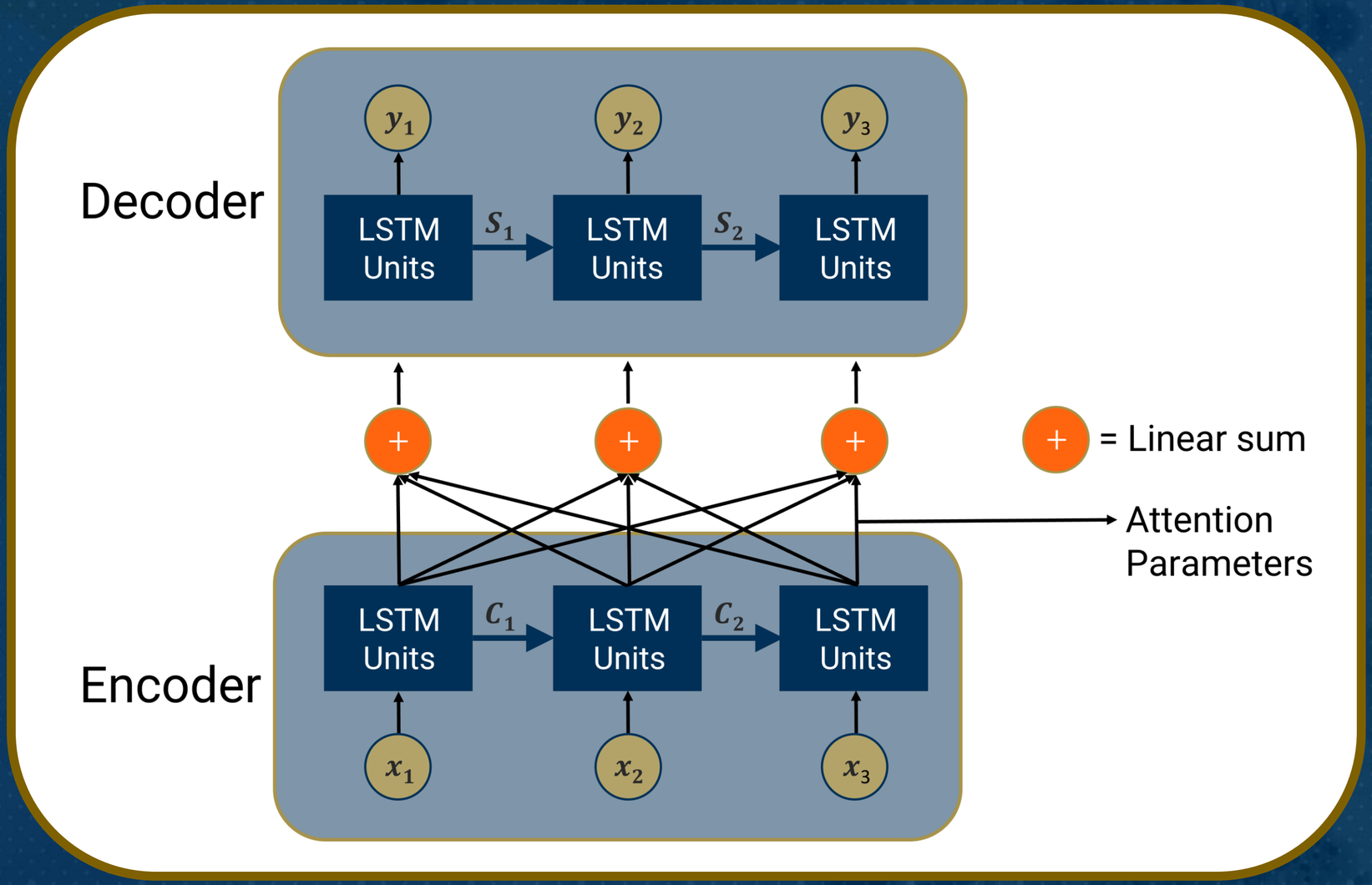
C: Final encoder state which is sent to a decoder

S_i: It is the decoder state

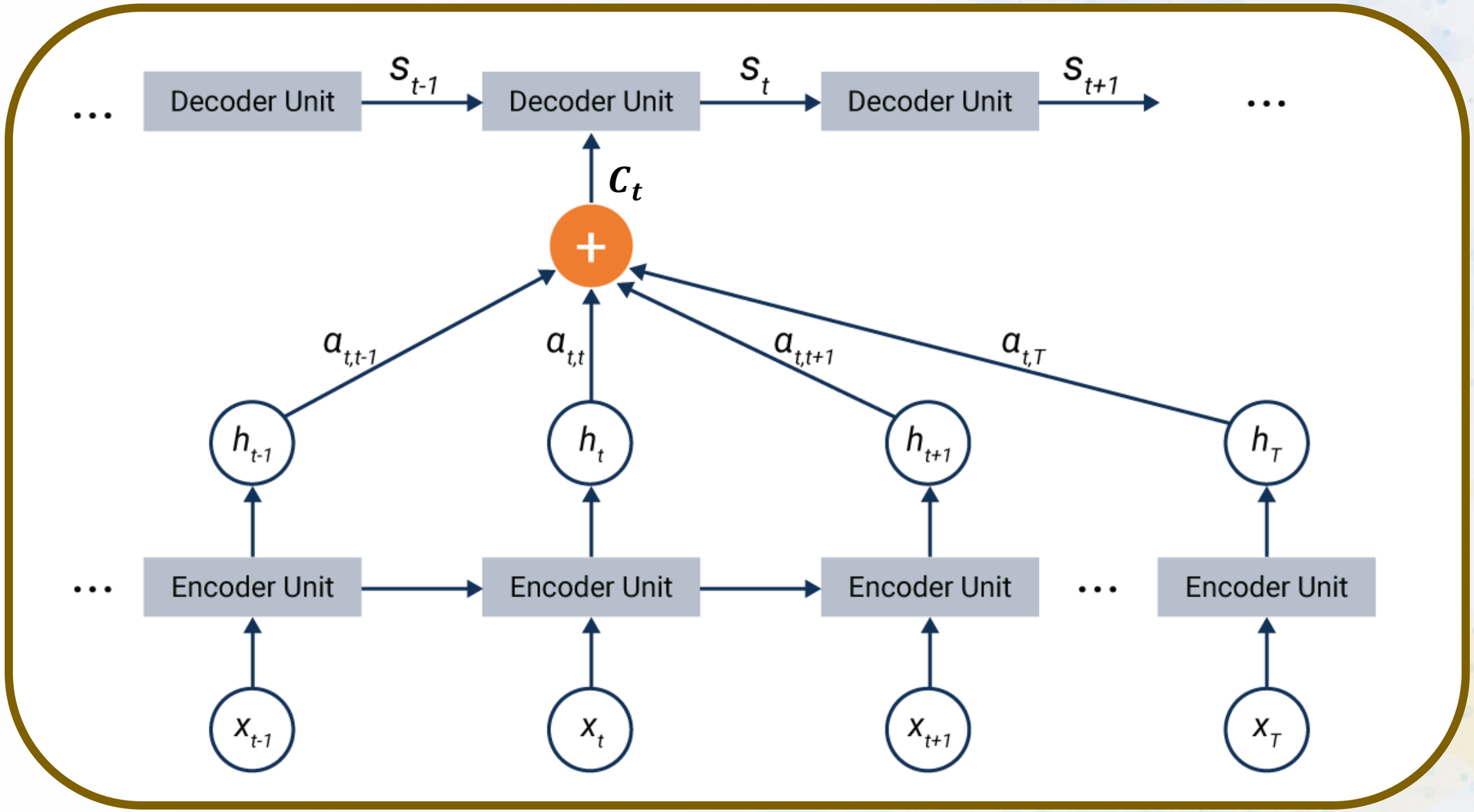
Why it is the Main Issue Here?



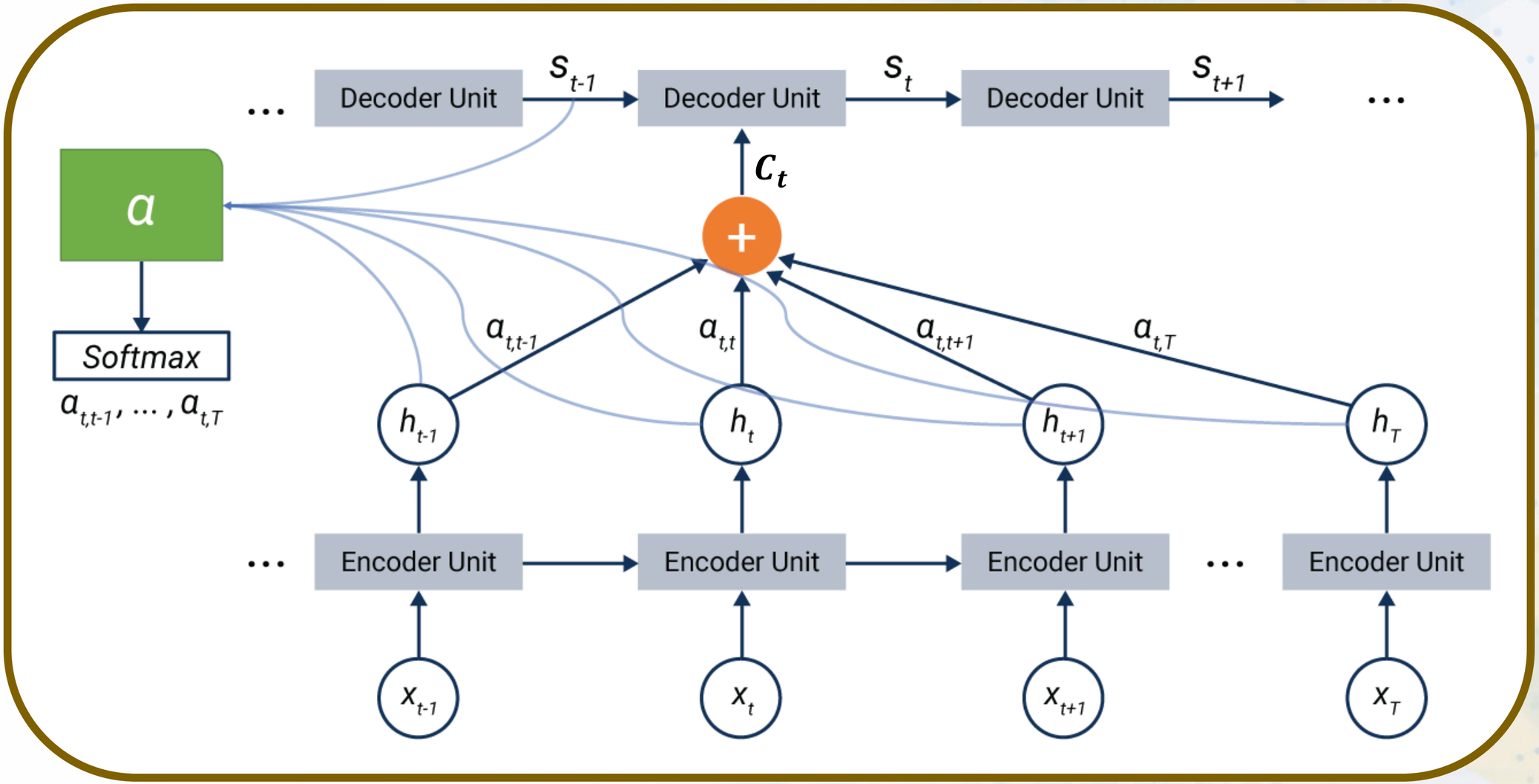
Attention Mechanism



We Need to Know the Weights Associated with Each Attention



We Need to Know the Weights Associated with Each Attention



Summary

- We learned about the Attention Mechanism
- Encoder and Decoder