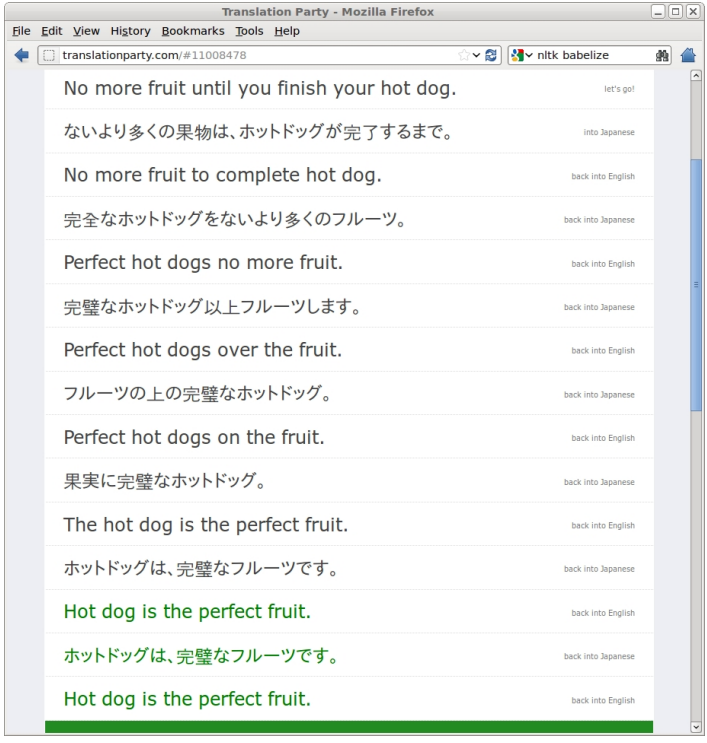
Semester end:

- RNNs and LSTMs (last week Monday)
- Machine translation (**Today**)
- Machine translation lab (Wednesday)
- Large language models and text generation (Friday and next week Monday)
- Ethical questions (next week Wednesday)
- Review for final exam (next week Friday)

Today:

- Background of the machine translation problem
- Derivation of the encoder-decoder model
- Specific tasks and difficulties of machine translation
- Evaluation of machine translation
- Social and ethical considerations

Diagrams from Jurafsky and Martin, *Speech and Language Processing*, 3rd ed draft (Jan 7, 2023), Chapters 9 & 13


Translation Party - Mozilla Firefox
File Edit View History Bookmarks Tools Help
translationparty.com/#11008478
nltk babelize
No more fruit until you finish your hot dog.
let's go!
ないより多くの果物は、ホットドッグが完了するまで。
into Japanese
No more fruit to complete hot dog.
back into English
完全なホットドッグをないより多くのフルーツ。
back into Japanese
Perfect hot dogs no more fruit.
back into English
完璧なホットドッグ以上フルーツします。
back into Japanese
Perfect hot dogs over the fruit.
back into English
フルーツの上の完璧なホットドッグ。
back into Japanese
Perfect hot dogs on the fruit.
back into English
果実に完璧なホットドッグ。
back into Japanese
The hot dog is the perfect fruit.
back into English
ホットドッグは、完璧なフルーツです。
back into Japanese
Hot dog is the perfect fruit.
back into English
ホットドッグは、完璧なフルーツです。
back into Japanese
Hot dog is the perfect fruit.
back into English

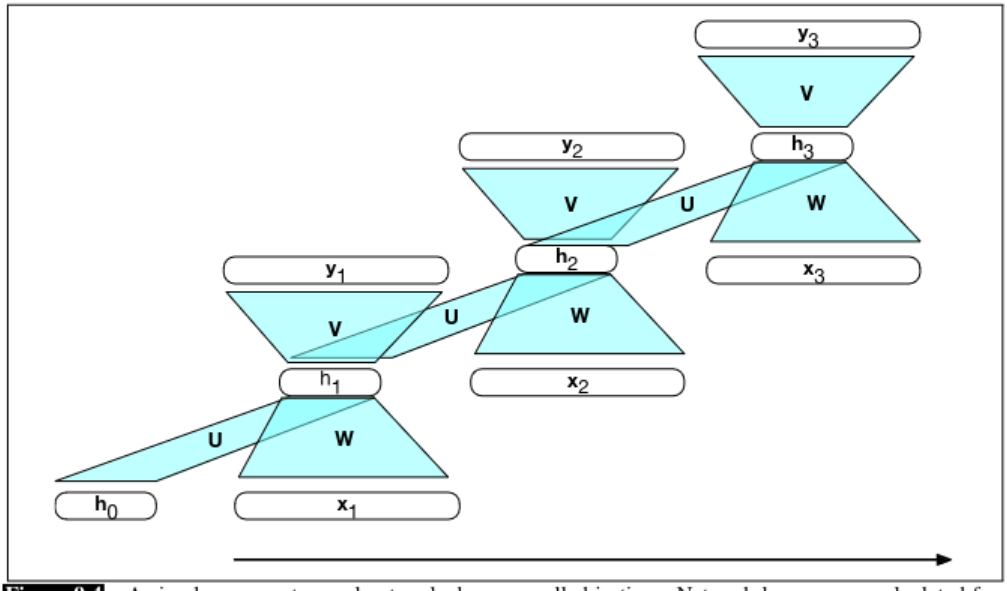


Figure 9.4 A simple recurrent neural network shown unrolled in time. Network layers are recalculated for each time step, while the weights **U**, **V** and **W** are shared across all time steps.

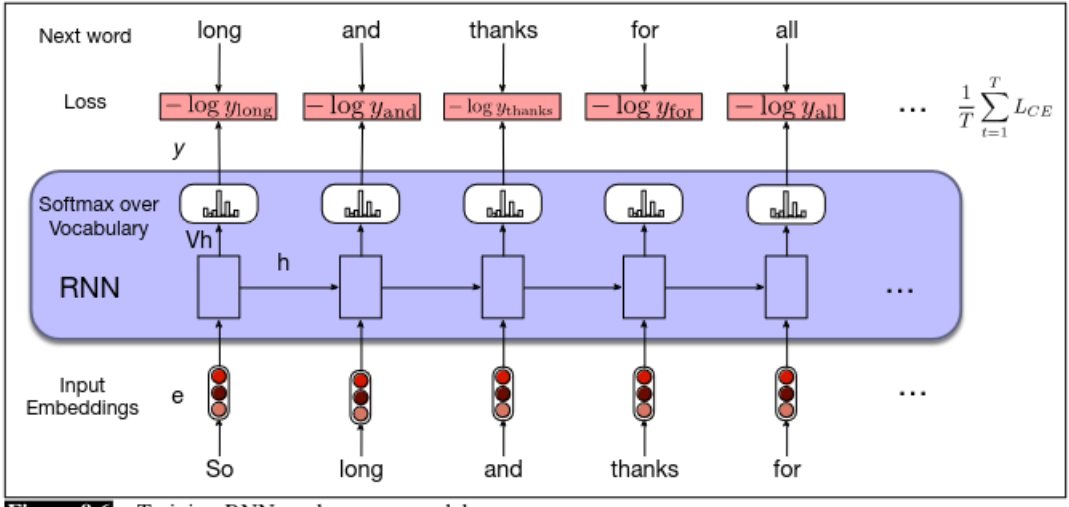


Figure 9.6 Training RNNs as language models.

Jurafsky and Martin, Sec 9.2, pg 7

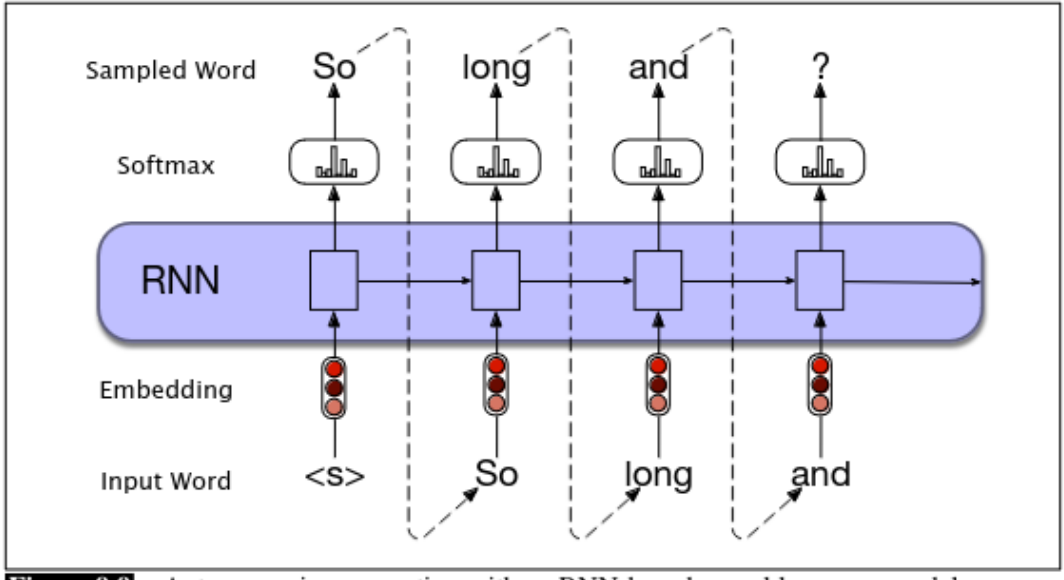


Figure 9.9 Autoregressive generation with an RNN-based neural language model.

Jurafsky and Martin, Sec 9.3, pg 11

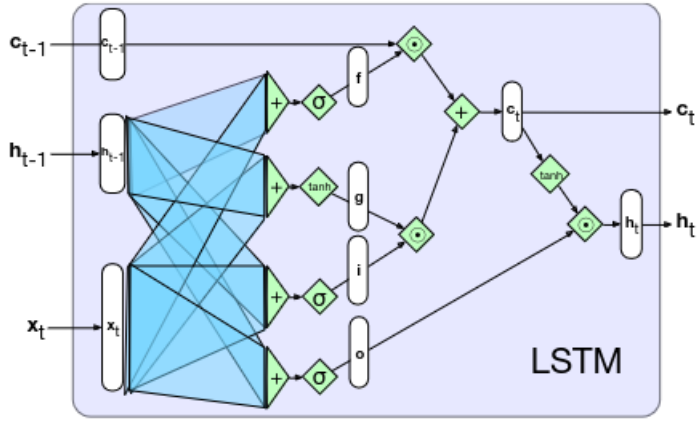


Figure 9.13 A single LSTM unit displayed as a computation graph. The inputs to each unit consists of the current input, x, the previous hidden state, h_{t-1}, and the previous context, c_{t-1}. The outputs are a new hidden state, h_t and an updated context, c_t.

Jurafsky and Martin, Sec 9.5, pg 16

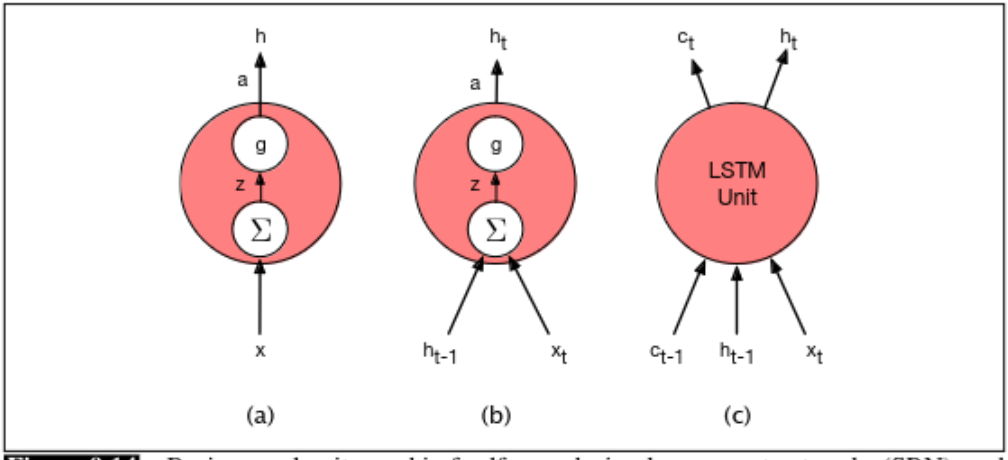


Figure 9.14 Basic neural units used in feedforward, simple recurrent networks (SRN), and long short-term memory (LSTM).

Jurafsky and Martin, Sec 9.5, pg 16

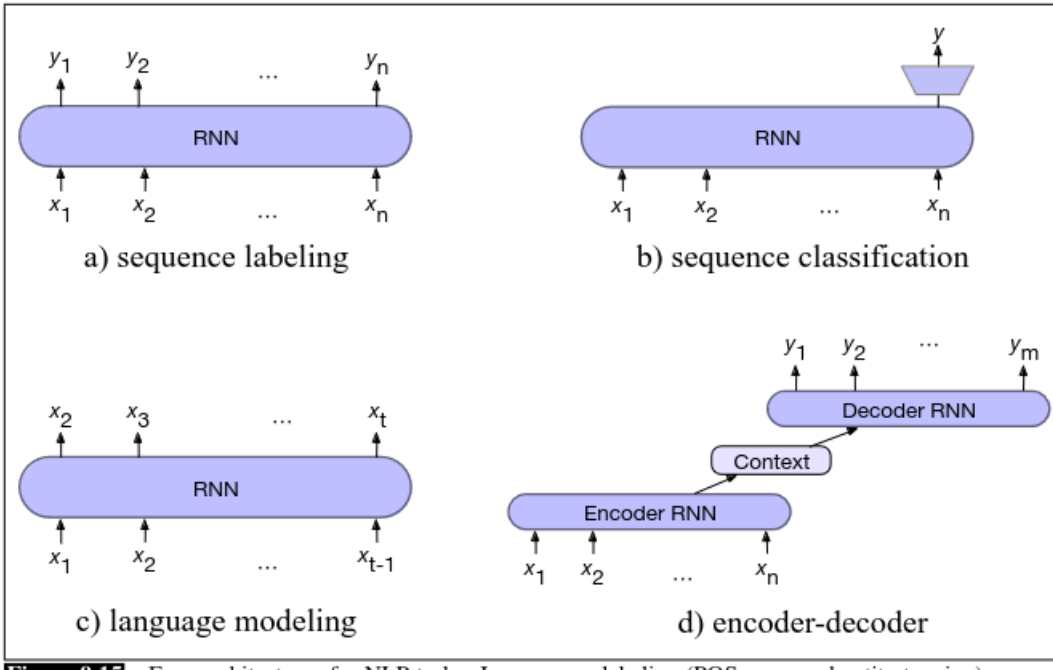


Figure 9.15 Four architectures for NLP tasks. In sequence labeling (POS or named entity tagging) we map

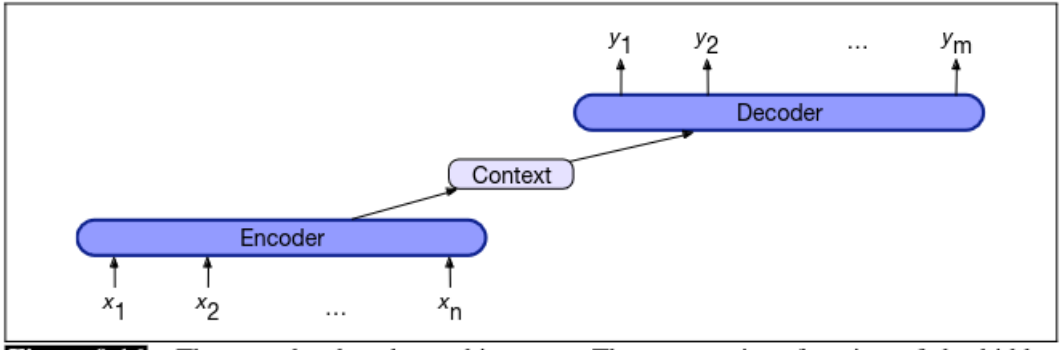


Figure 9.16 The encoder-decoder architecture. The context is a function of the hidden representations of the input, and may be used by the decoder in a variety of ways.

Jurafsky and Martin, Sec 9.7, pg 18

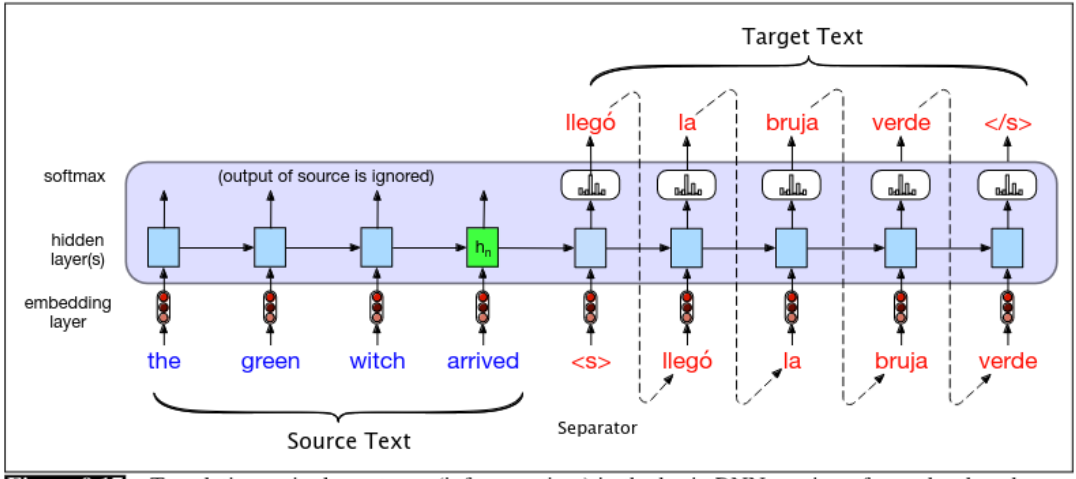
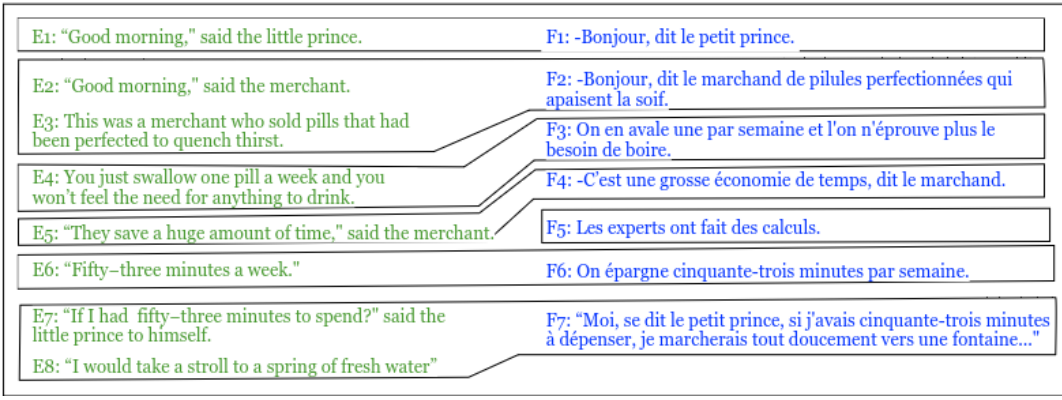


Figure 9.17 Translating a single sentence (inference time) in the basic RNN version of encoder-decoder approach to machine translation. Source and target sentences are concatenated with a separator token in between, and the decoder uses context information from the encoder's last hidden state.

Jurafsky and Martin, Sec 9.8, pg 22



Usually dynamic programming is used as the alignment algorithm (Gale and Church, 1993), in a simple extension of the minimum edit distance algorithm we introduced in Chapter 2.

Jurafsky and Martin, Sec 13.2 (as 10.2), pg 9

Coming up:

- Read J&M chapter 13 (Mon, Nov 27)
- Take the machine translation quiz (Tues, Nov 28)
- Finish stylometry assignment (Fri, Dec 1)
- Do word2Vec programming assignment (Wed, Dec 6)

Remaining things for this semester (not yet posted to Canvas, due dates not set)

- A reading on large language models and/or text generation (not from J&M)
- Another quizz or two
- A short reflection or two on ethical or social questions