



Adapted from material by Thang Luong

Machine Translation

Computational Linguistics: Jordan Boyd-Graber
University of Maryland

NEURAL MODELS

What makes neural models different?

- Sequence-to-sequence: train directly on sentence pairs
- No intermediate steps (e.g., alignment): generalizable
- Fluent!

What makes neural models different?

- Sequence-to-sequence: train directly on sentence pairs
- No intermediate steps (e.g., alignment): generalizable
- Fluent! (Looks better, hard to diagnose errors)

What makes neural models different?

- Sequence-to-sequence: train directly on sentence pairs
- No intermediate steps (e.g., alignment): generalizable
- Fluent! (Looks better, hard to diagnose errors)
- More popular, seems to work better (with enough data)

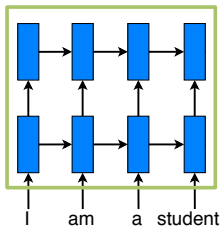
RNNs to the rescue

I am a student

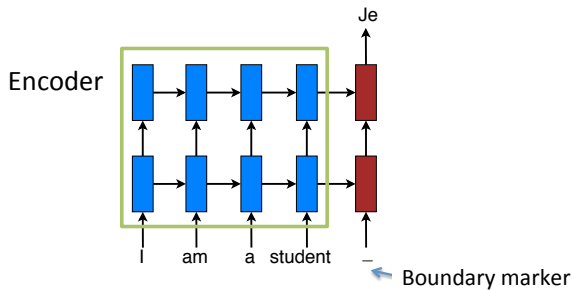
Je suis étudiant

RNNs to the rescue

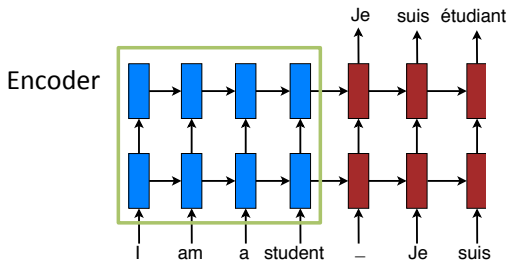
Encoder



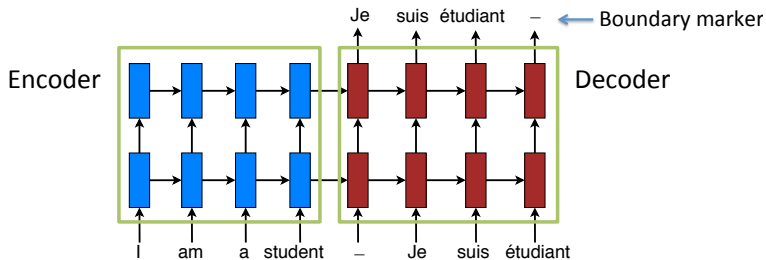
RNNs to the rescue



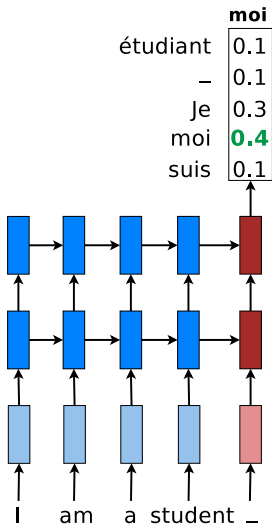
RNNs to the rescue



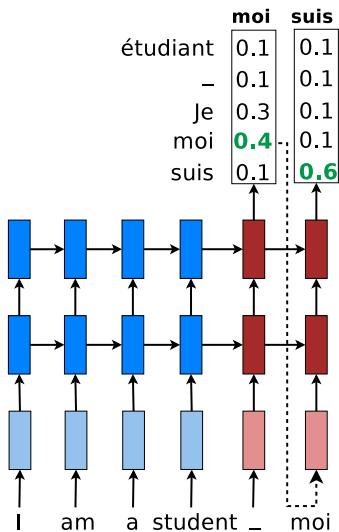
RNNs to the rescue



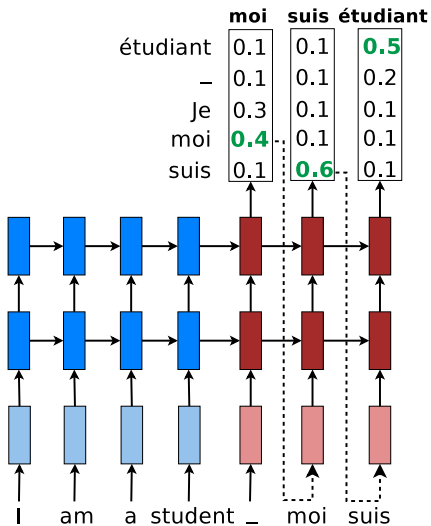
Test time ...



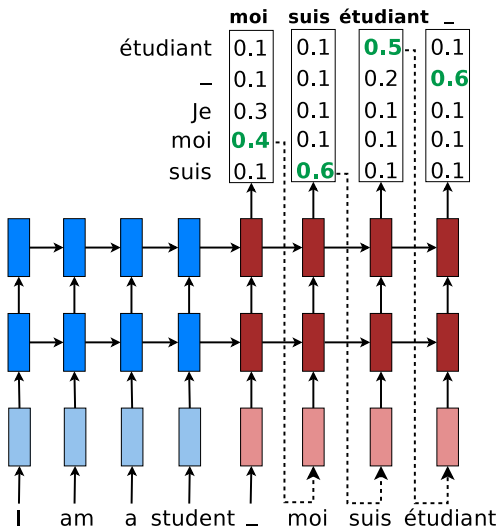
Test time ...



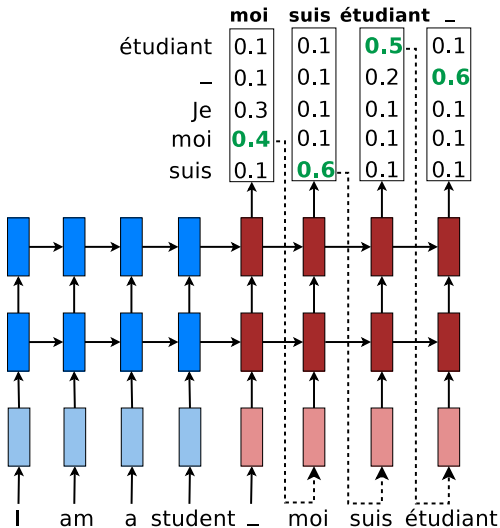
Test time ...



Test time ...



Test time ...



Problems

- Out of vocabulary words
- Complicated morphology
- Long sentences

Problems

- Out of vocabulary words: copying/transliteration
- Complicated morphology
- Long sentences

Problems

- Out of vocabulary words: copying/transliteration
- Complicated morphology: character-level models
- Long sentences

Problems

- Out of vocabulary words: copying/transliteration
- Complicated morphology: character-level models
- Long sentences: attention

Attention

知 识 就 是 力 量 <end>

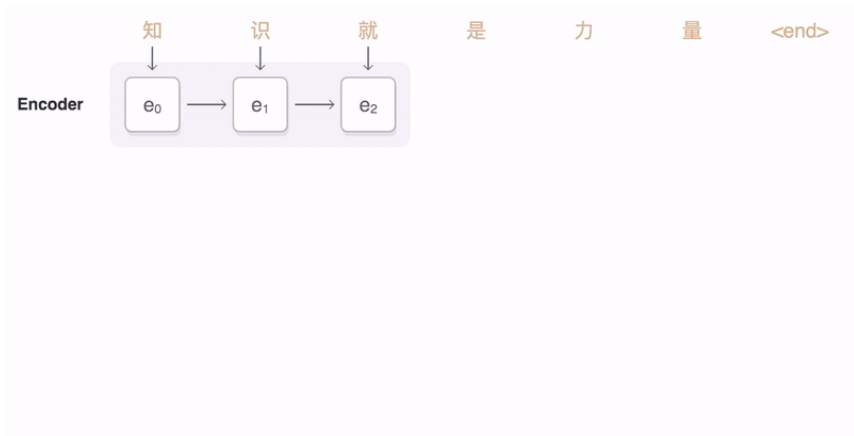
Attention



Attention



Attention



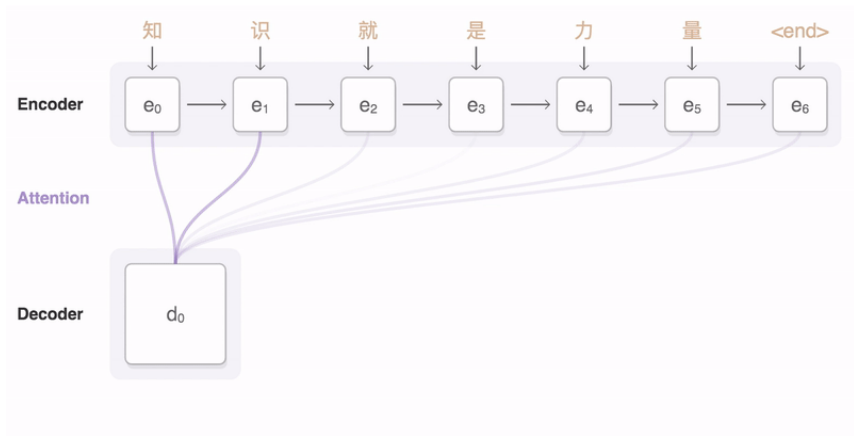
Attention



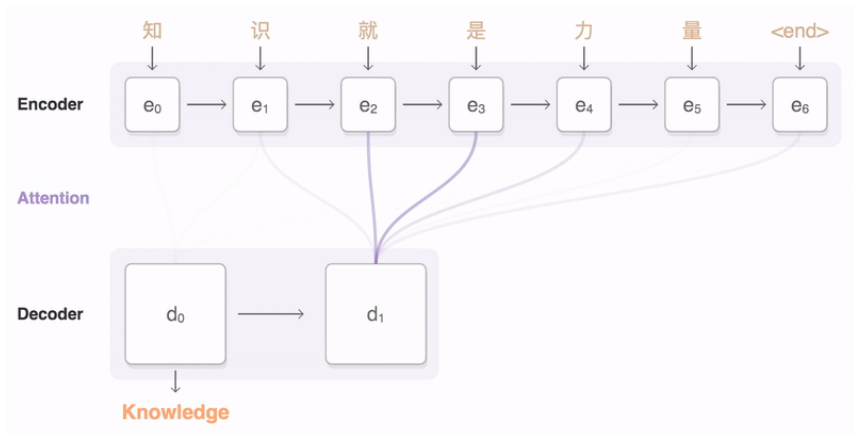
Attention



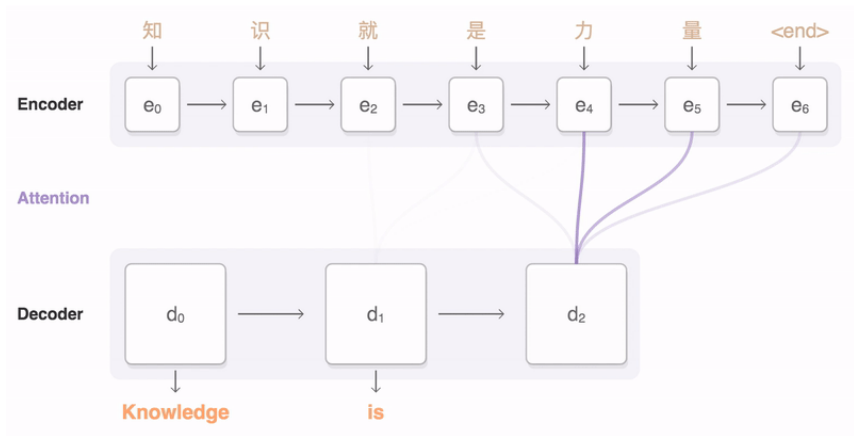
Attention



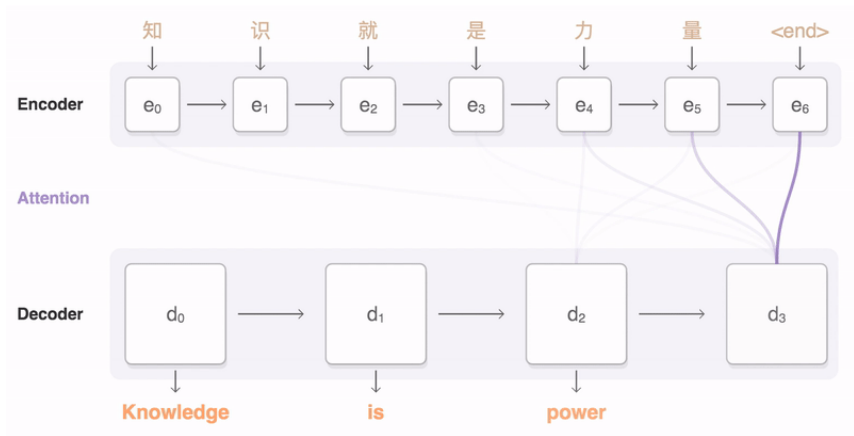
Attention



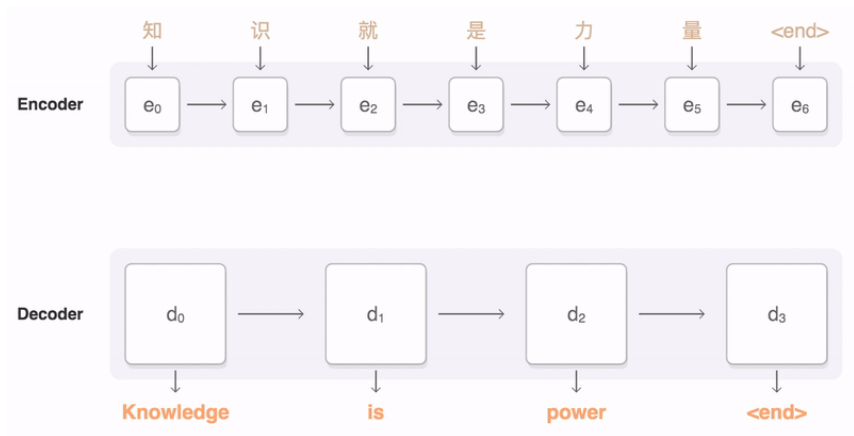
Attention



Attention



Attention



Beyond simple models

- Active area of research
- Implementations in all frameworks
- Reinforcement learning is huge component
- Low data