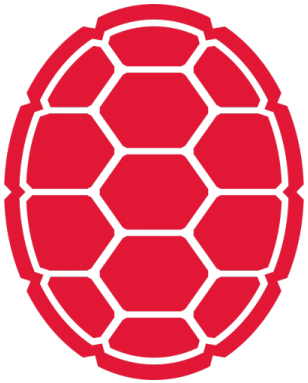


Machine Translation

Computational Linguistics: Jordan Boyd-Graber
University of Maryland

NEURAL MODELS

Adapted from material by Thang Luong

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)

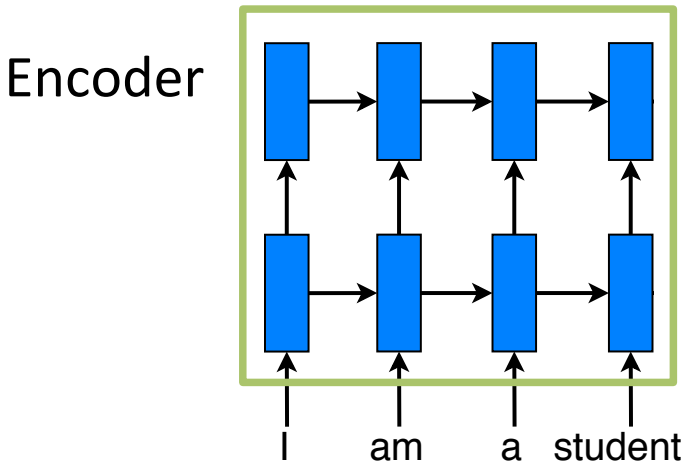
What makes neural models different?

- 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)

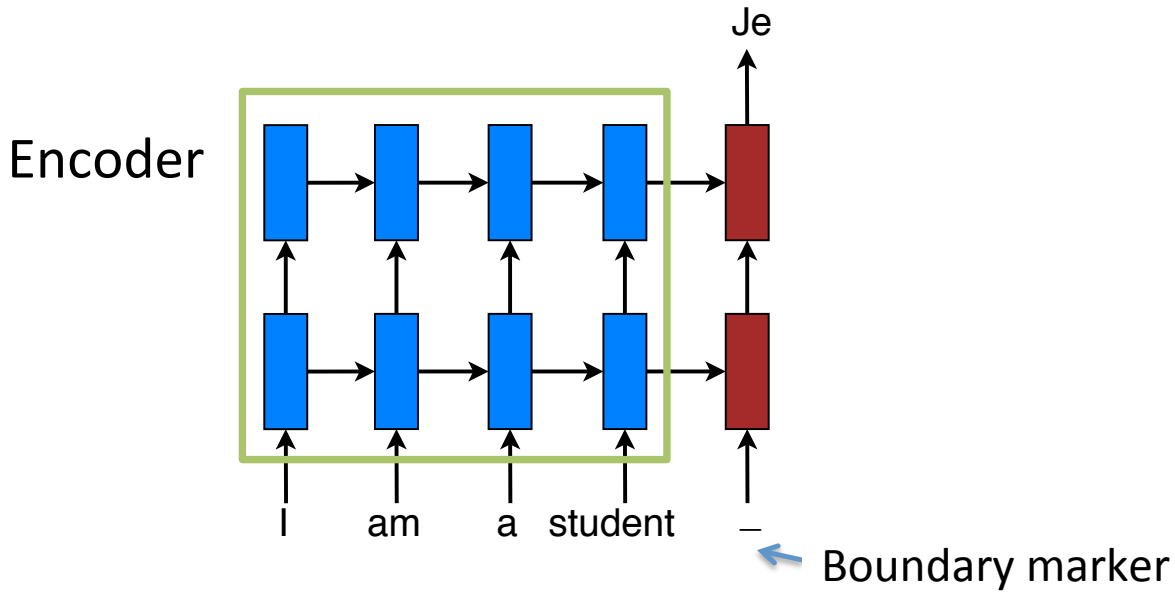
RNNs to the rescue

I am a student Je suis étudiant

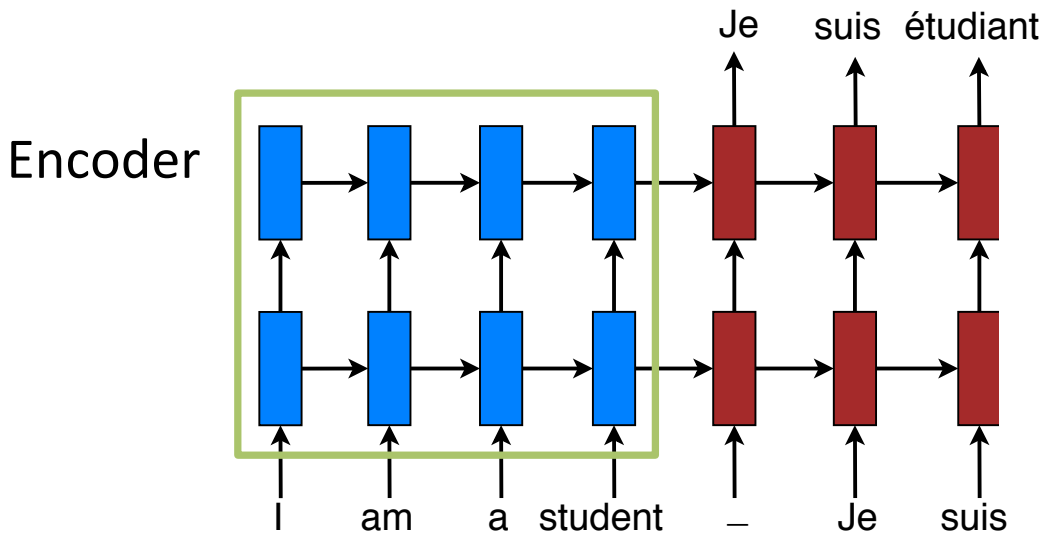
RNNs to the rescue



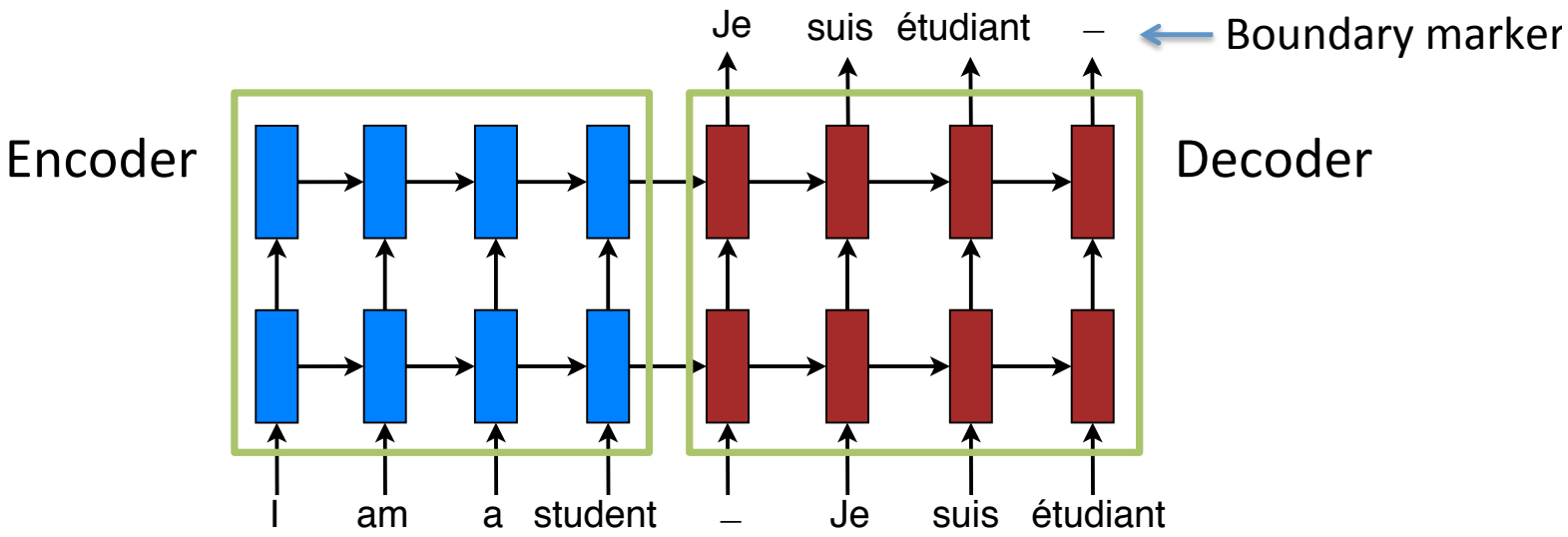
RNNs to the rescue



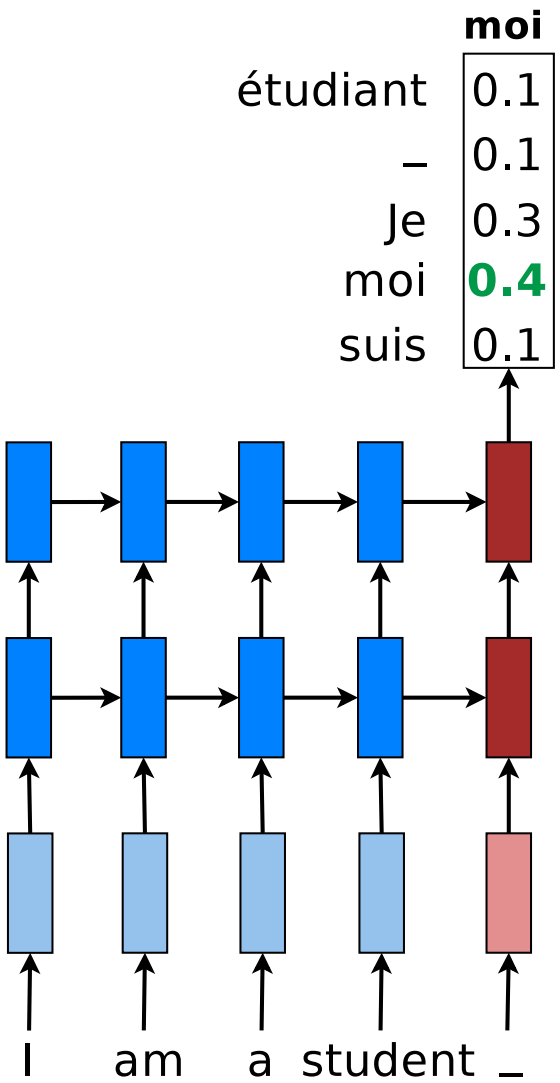
RNNs to the rescue



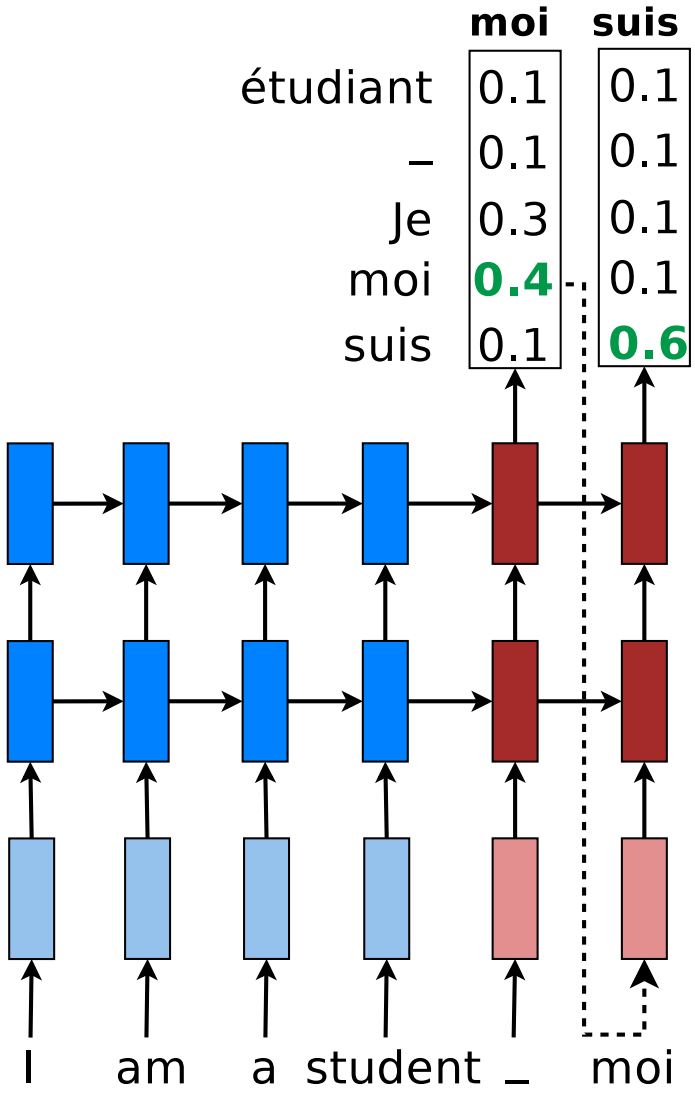
RNNs to the rescue



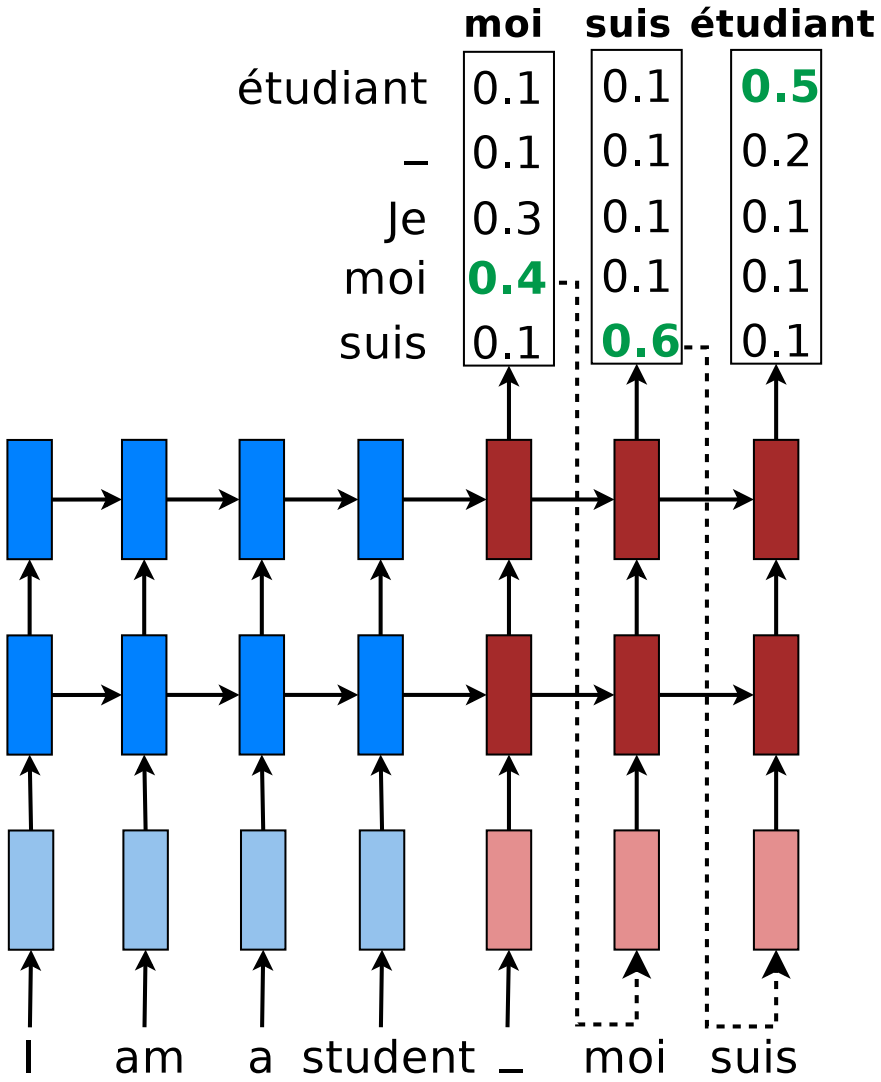
Test time . . .



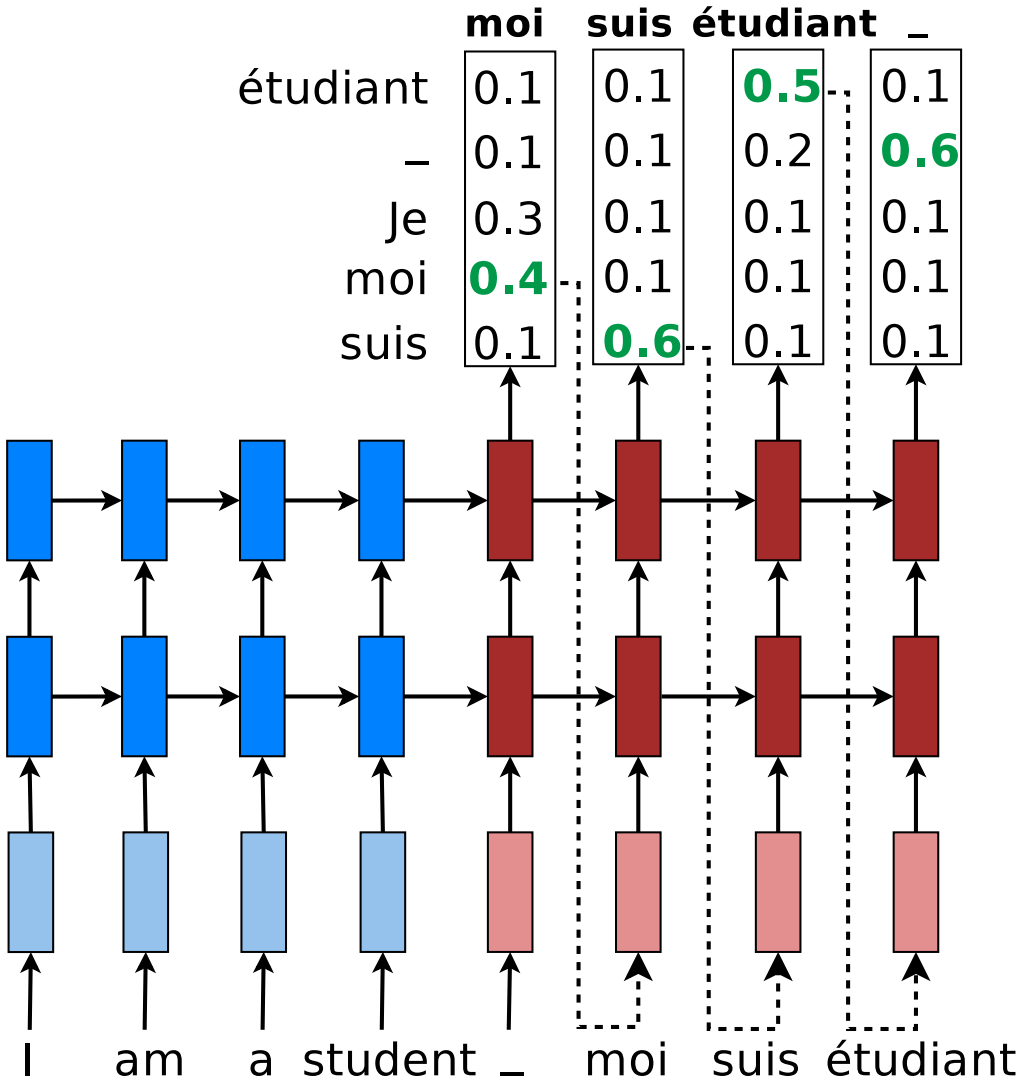
Test time . . .


moi
suis
étudiant 0.1 0.1
_ 0.1 0.1
Je 0.3 0.1
moi 0.4 0.1
suis 0.1 0.6
I
am
a
student
_
moi

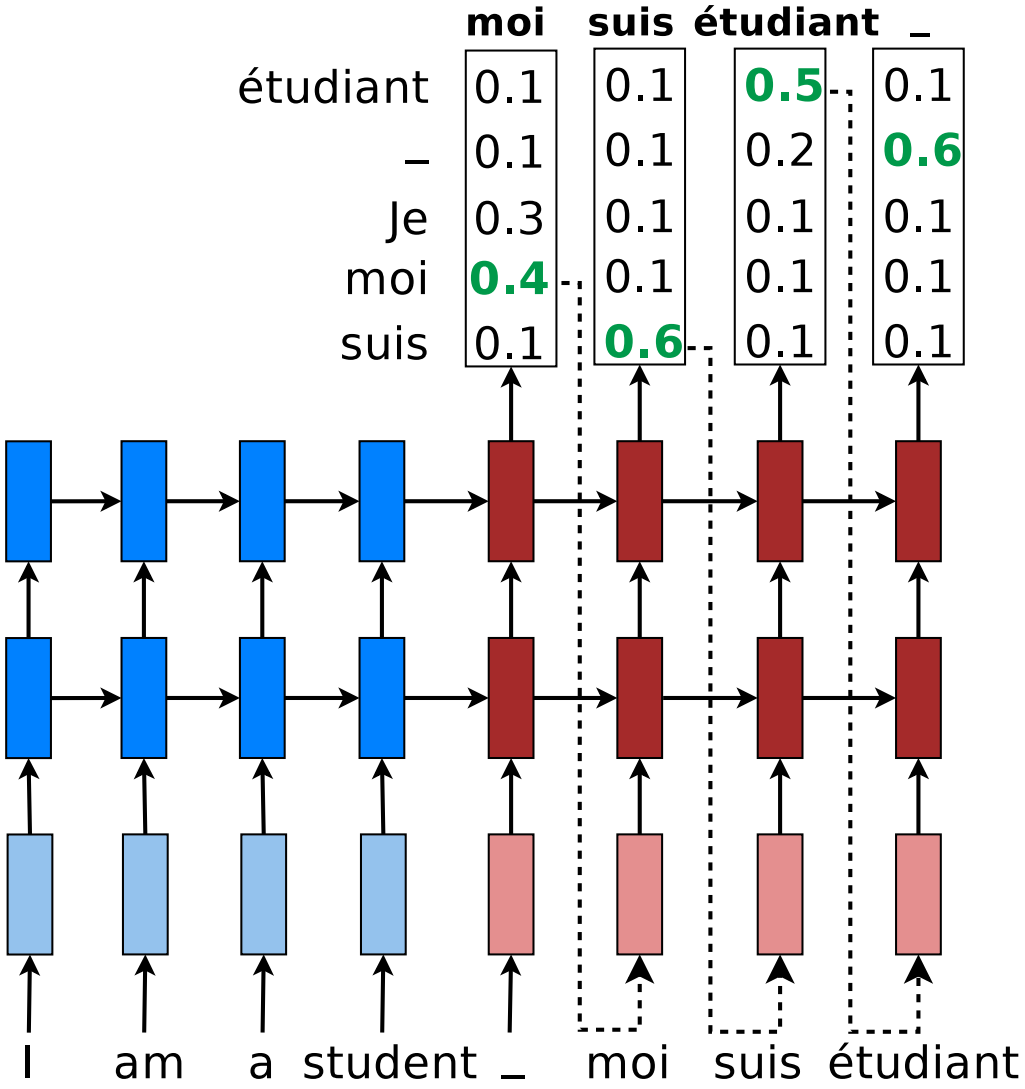
Test time . . .


moi
suis
étudiant
étudiant 0.1 0.1 0.5
_ 0.1 0.1 0.2
Je 0.3 0.1 0.1
moi 0.4 0.1 0.1
suis 0.1 0.6 0.1
I am a student _ moi suis

Test time . . .


moi suis étudiant _
étudiant 0.1 0.1 0.5 0.1
_ 0.1 0.1 0.2 0.6
Je 0.3 0.1 0.1 0.1
moi 0.4 0.1 0.1 0.1
suis 0.1 0.6 0.1 0.1
I am a student _ moi suis étudiant

Test time . . .


moi suis étudiant _
étudiant 0.1 0.1 0.5 0.1
_ 0.1 0.1 0.2 0.6
Je 0.3 0.1 0.1 0.1
moi 0.4 0.1 0.1 0.1
suis 0.1 0.6 0.1 0.1
I am a student _ moi suis étudiant

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

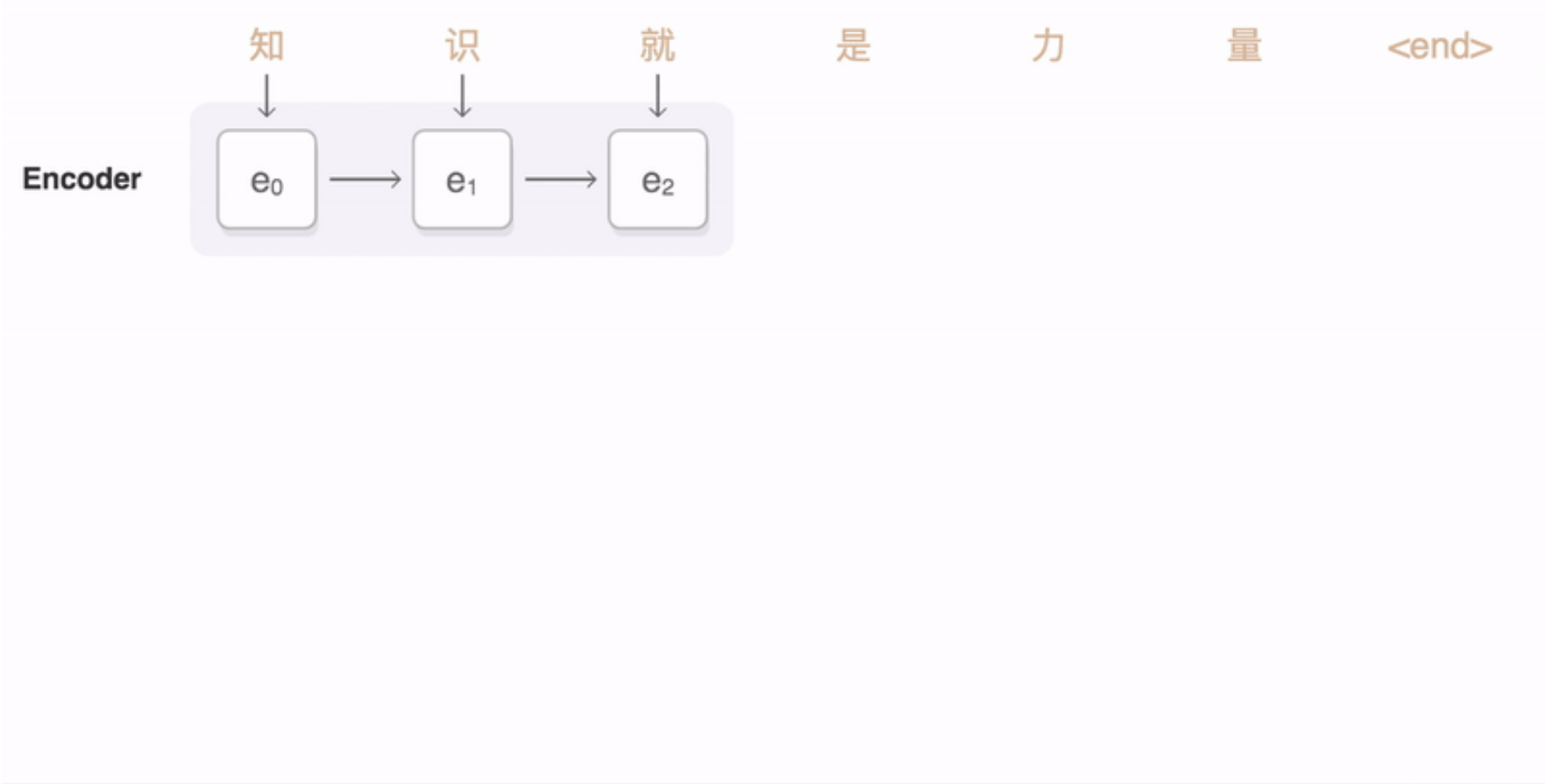
- Out of vocabulary words: copying/transliteration
- Complicated morphology: character-level models
- Long sentences: attention

Attention

Attention

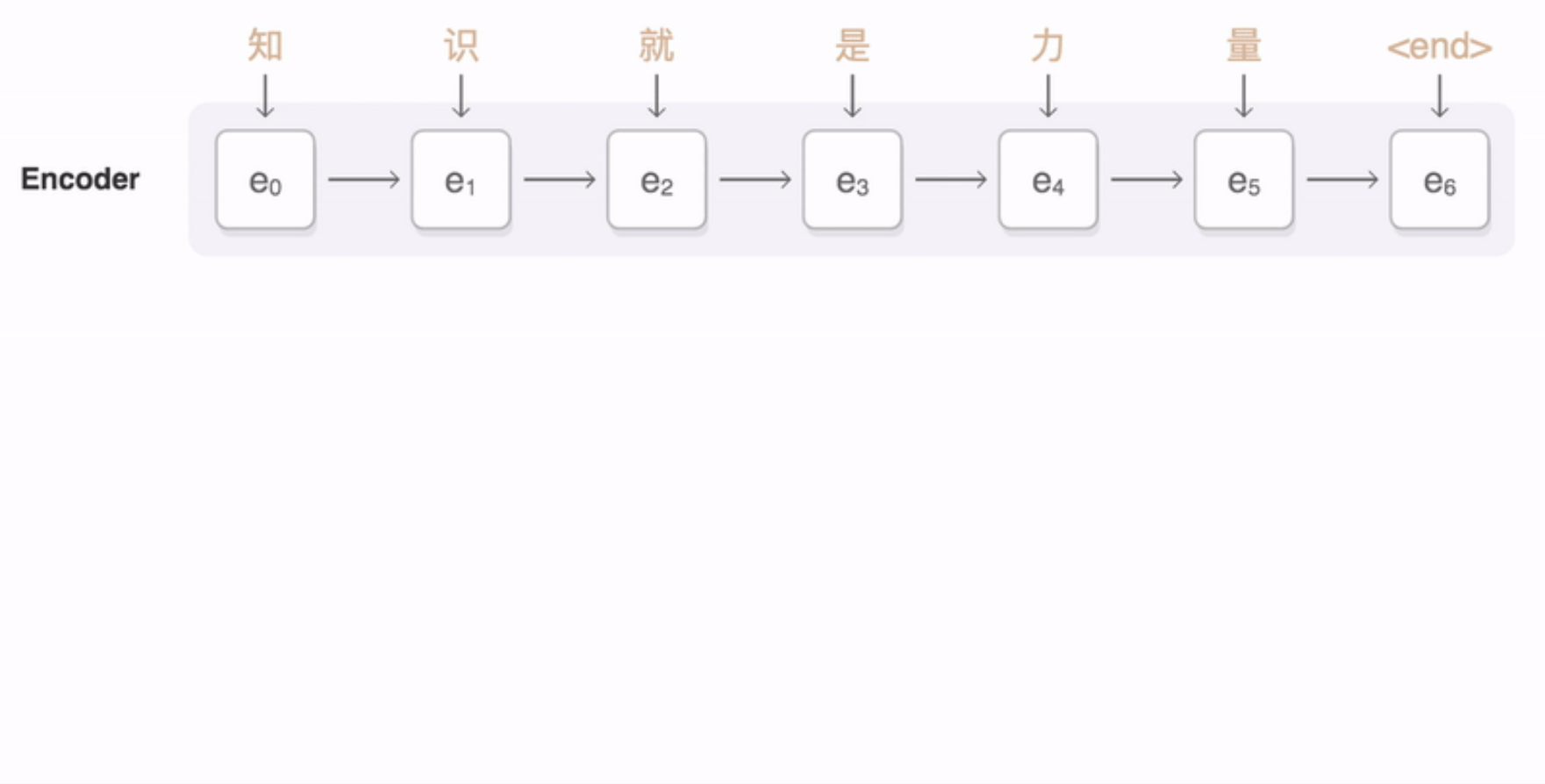
Attention

Attention

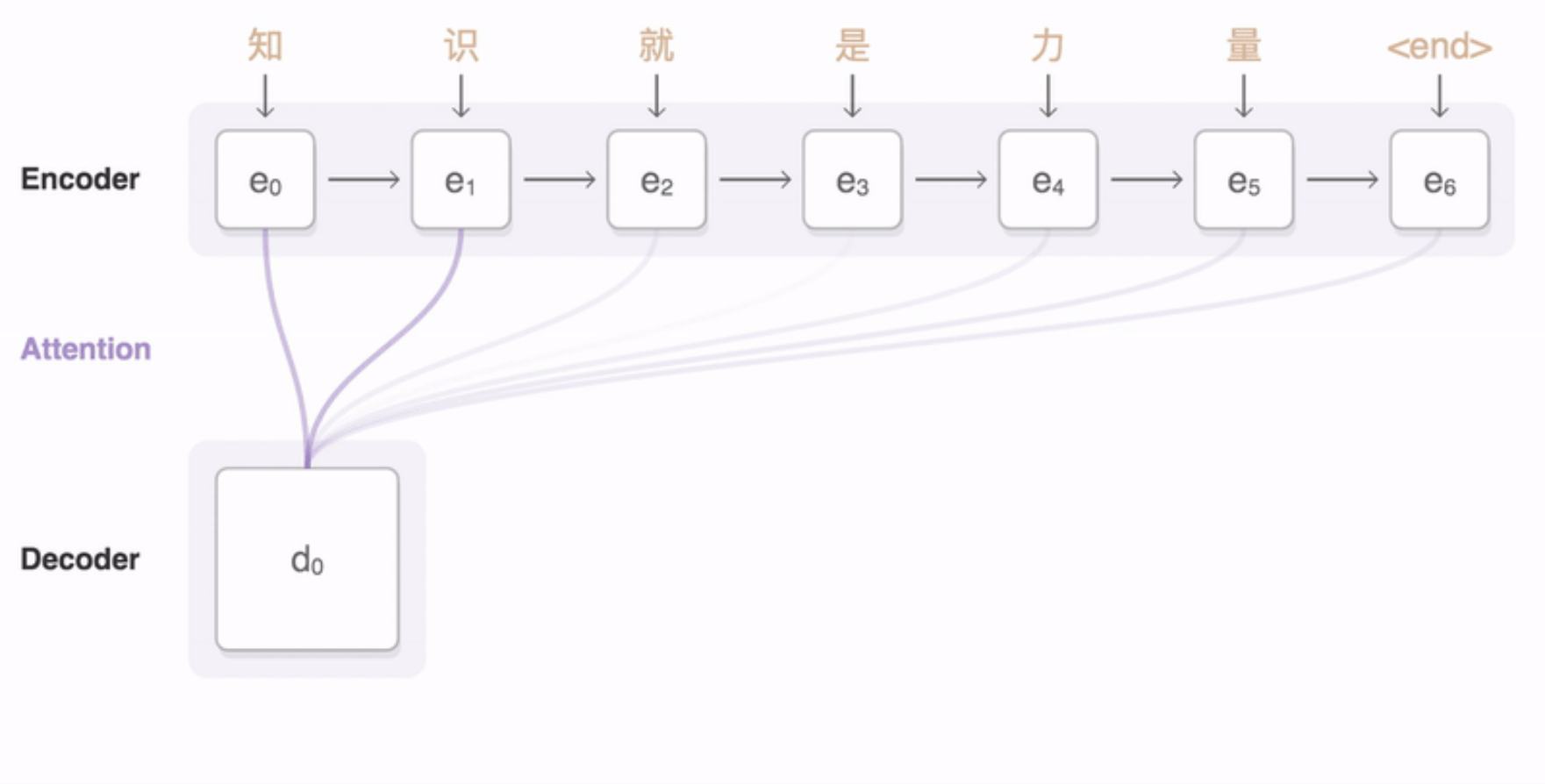


Attention

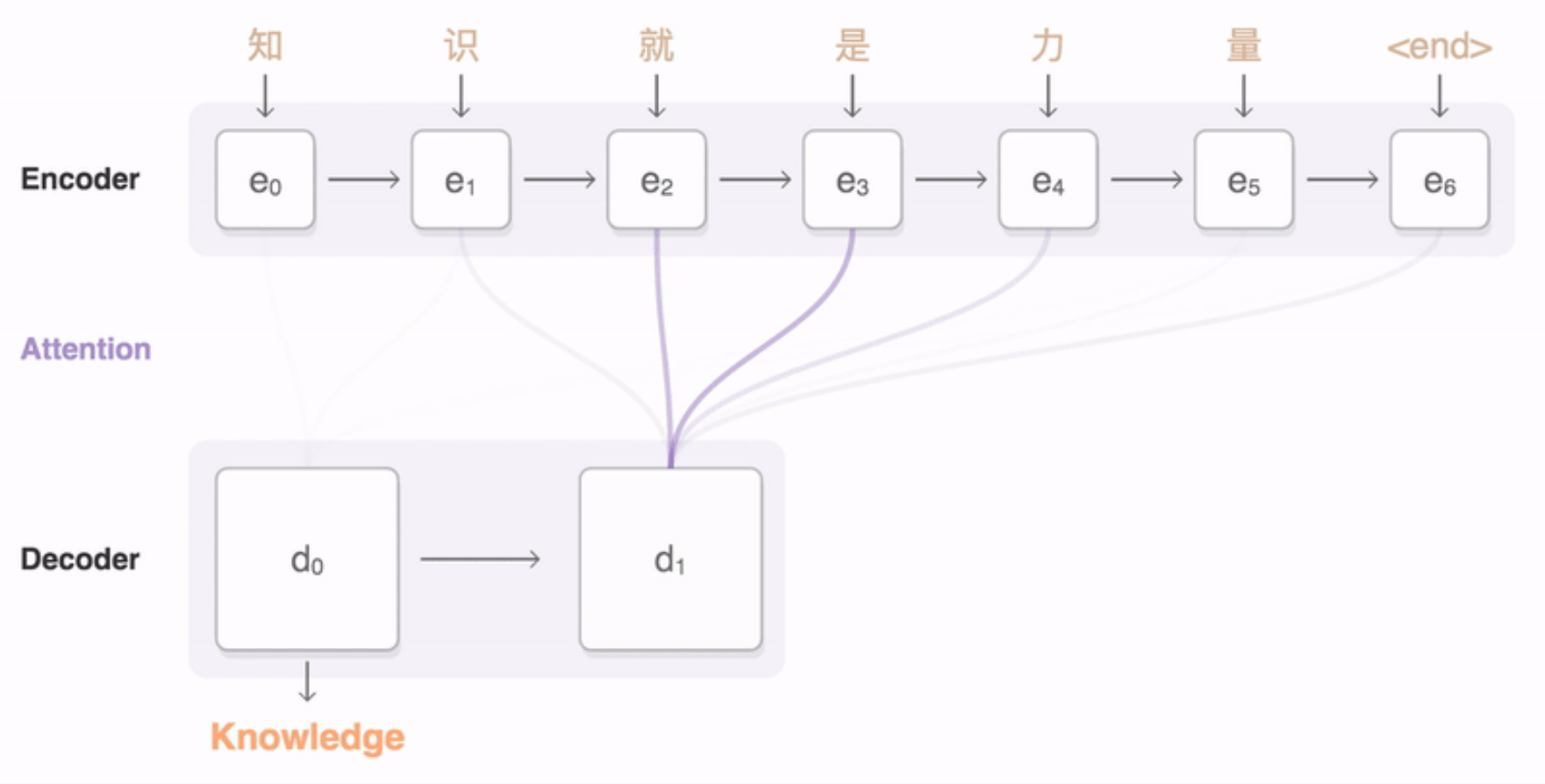
Attention



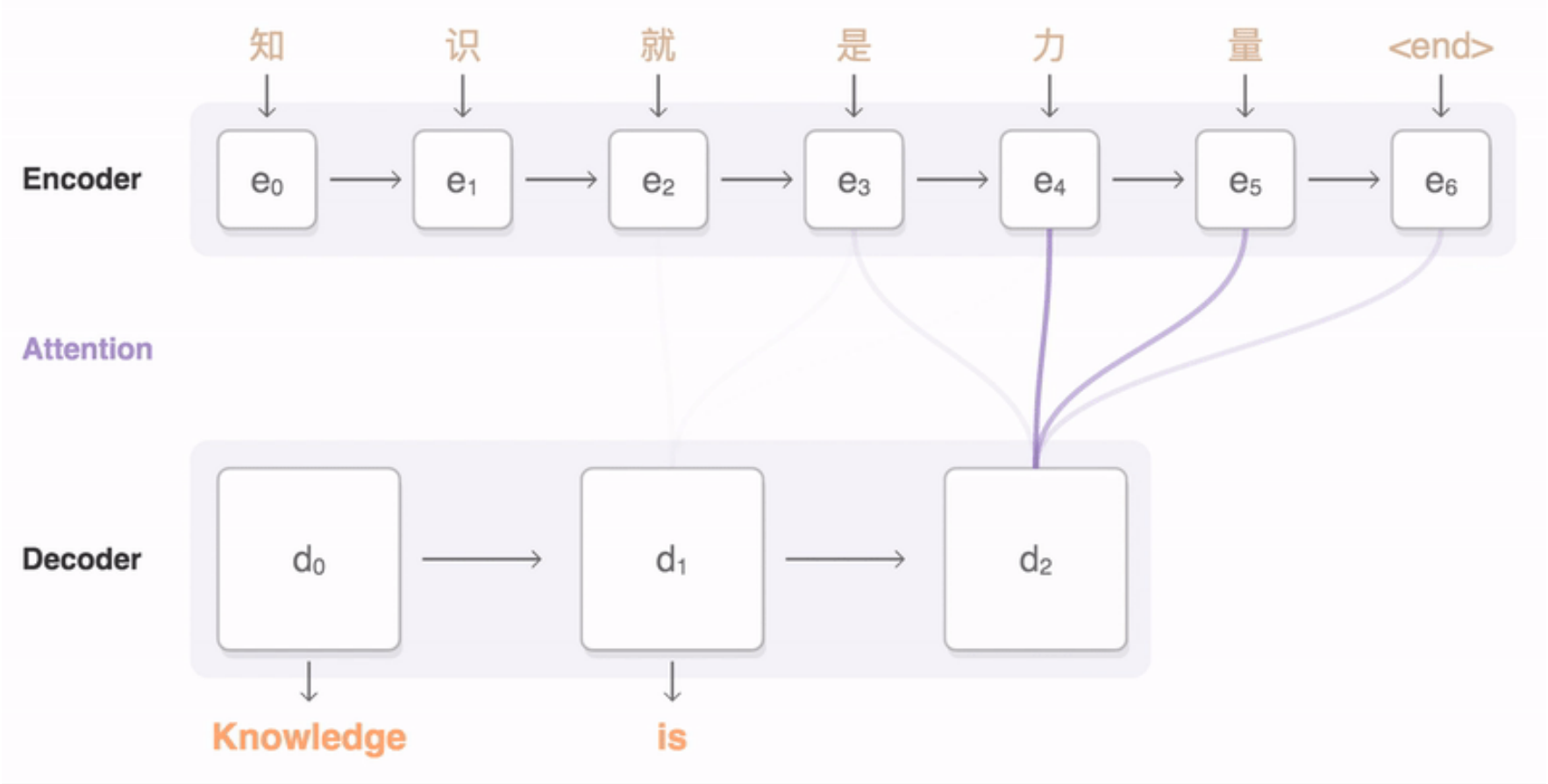
Attention



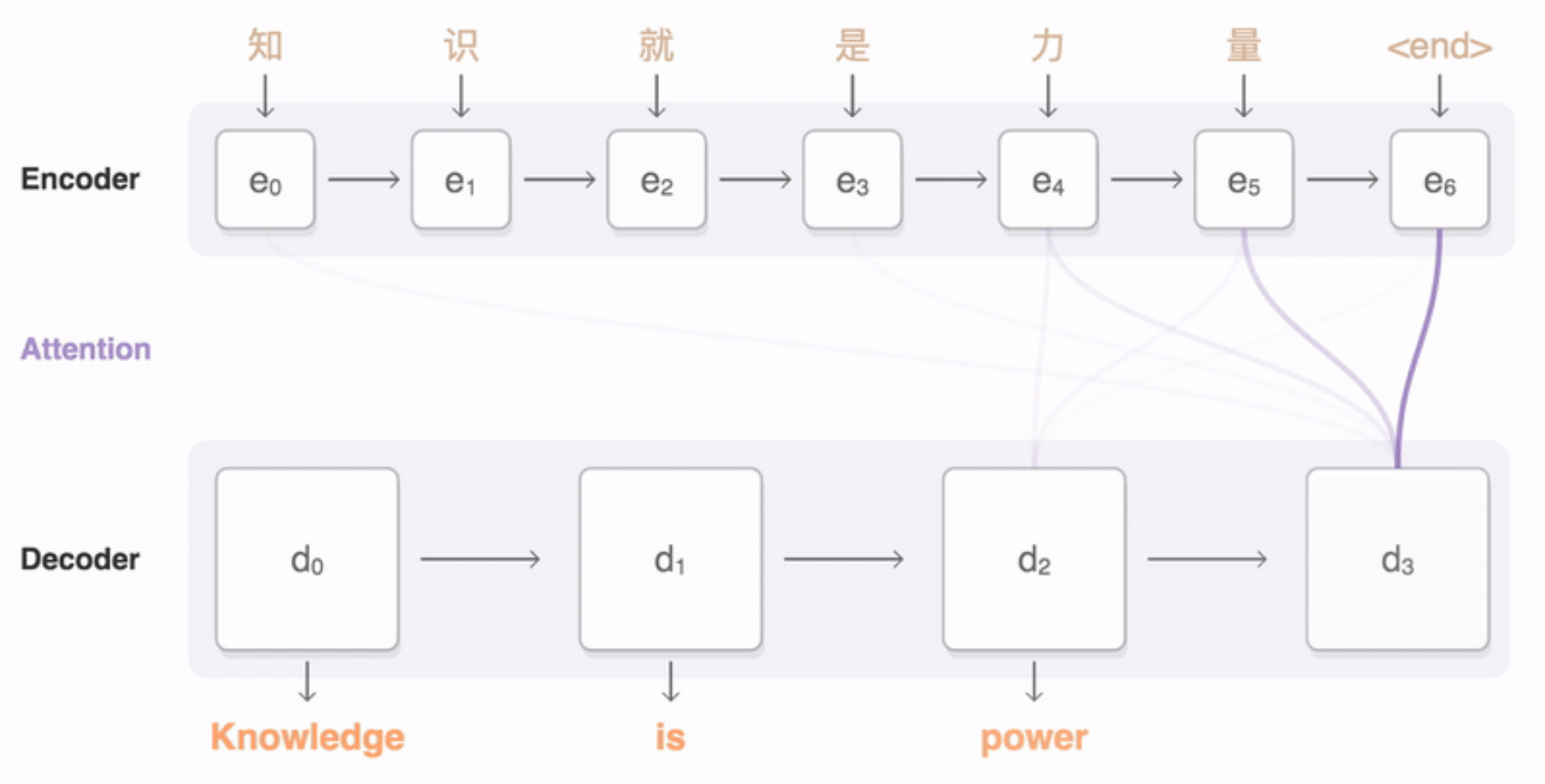
Attention


知
识
就
是
力
量
<end>
Encoder
e0
e1
e2
e3
e4
e5
e6
Attention
Decoder
d0
d1
Knowledge

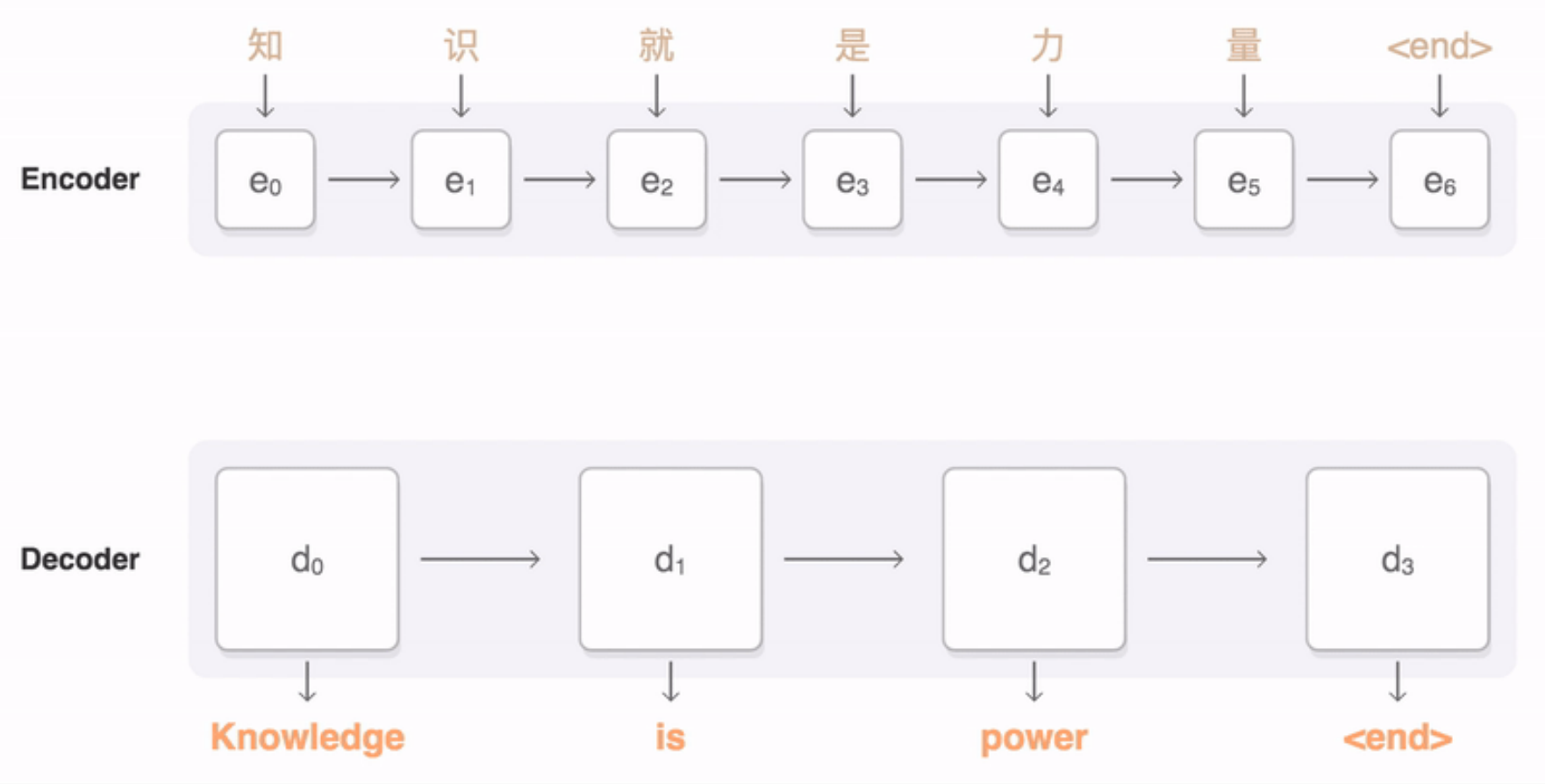
Attention


知
识
就
是
力
量
<end>
Encoder
e0
e1
e2
e3
e4
e5
e6
Attention
Decoder
d0
d1
d2
Knowledge
is

Attention


知
识
就
是
力
量
<end>
Encoder
e0
e1
e2
e3
e4
e5
e6
Attention
Decoder
d0
d1
d2
d3
Knowledge
is
power

Attention


知
识
就
是
力
量
<end>
Encoder
e0
e1
e2
e3
e4
e5
e6
Decoder
d0
d1
d2
d3
Knowledge
is
power
<end>

Beyond simple models

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