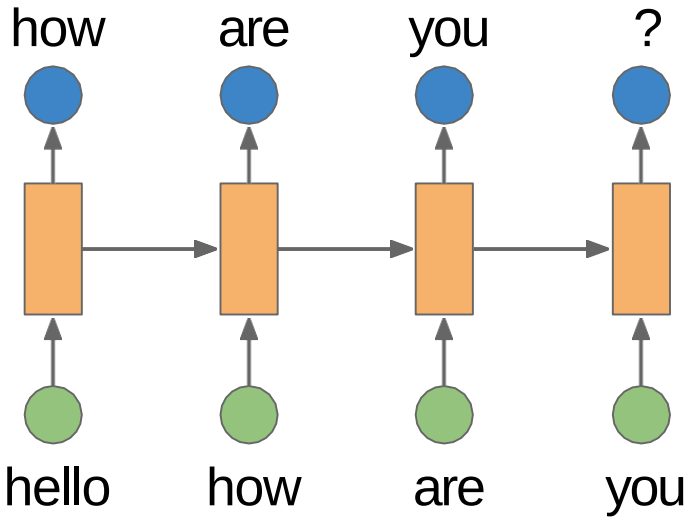
Sequence to Sequence Models

Lecture # 7

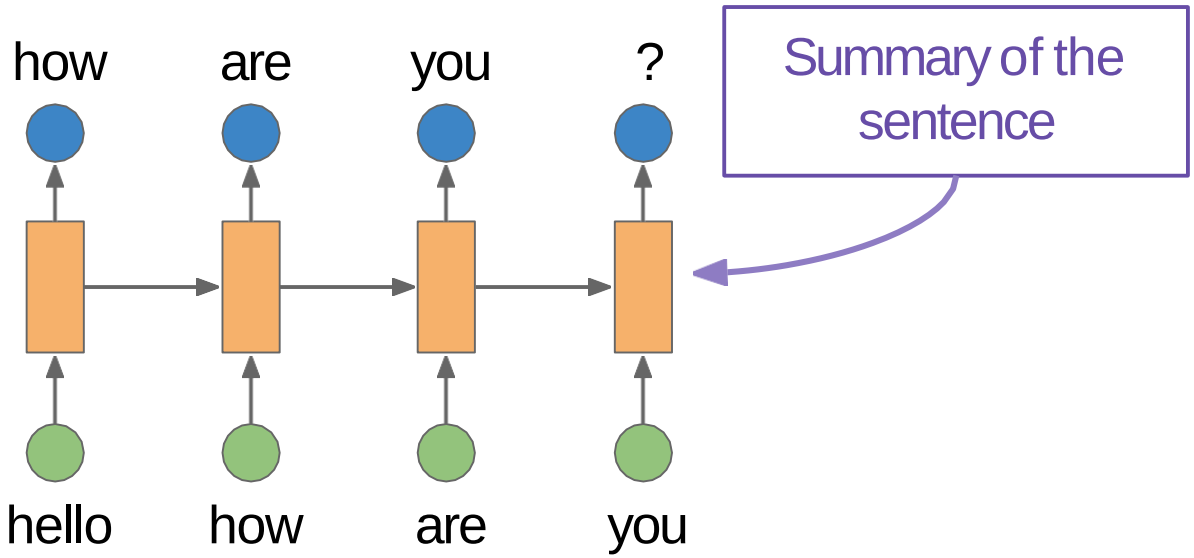
Recap

- RNN predicts a word based on the context
- A context vector at time t can be seen as a summary of previous words



Recap

- RNN predicts a word based on the context
- A context vector at time t can be seen as a summary of previous words



Sequence to Sequence Models

Seq2Seq models take RNN a step further

Sequence to Sequence Models

Seq2Seq models take RNN a step further

Seq2Seq models take a sequence and predict another sequence

Sequence to Sequence Models

Seq2Seq models take RNN a step further

Seq2Seq models take a sequence and predict another sequence

- Translation: Parallel source and target sentences
  - sequence of words → sequence of words
- Speech Recognition: Audio waves to transcription
  - sequence of audio signals → sequence of words
- Image Captioning: Images to text sequence
  - "sequence" of pixels → sequence of words

Sequence to Sequence Models

Intuitively, the **seq2seq** model learns to:

- Read a source sequence completely
- Predict a target sequence

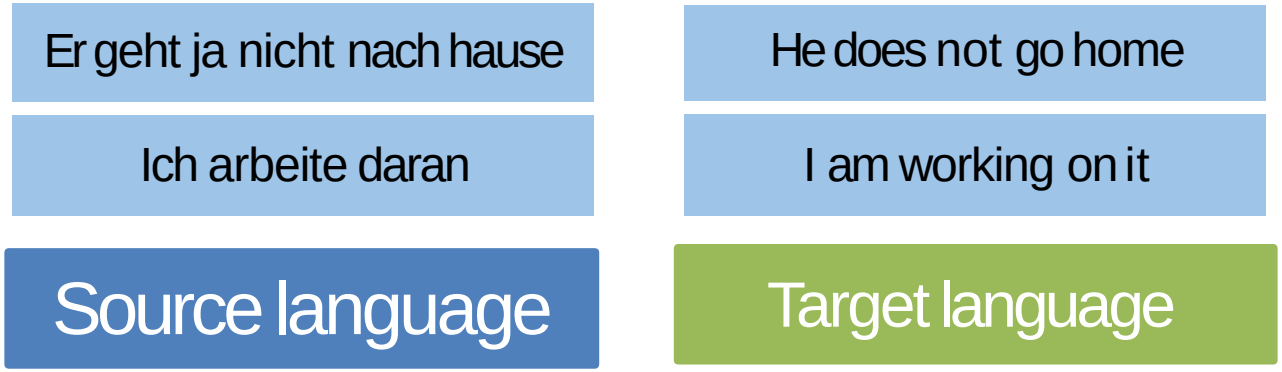
Sequence to Sequence Models

Different from

- *Per timestep prediction* using RNN, here both source and target are sequences
- *Sequence generation* using RNN, both source and target are from different worlds

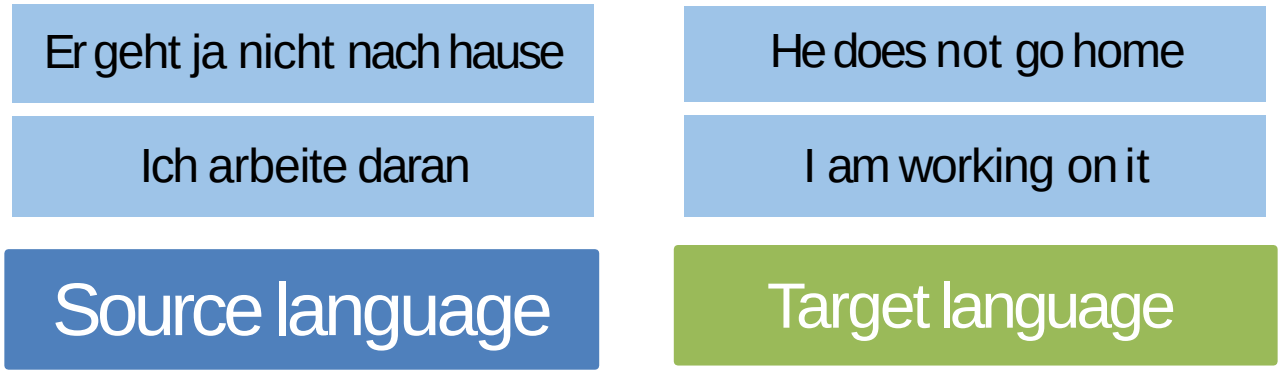
Sequence to Sequence Models

- Machine translation as an example
  - Pairs of sentences in two languages



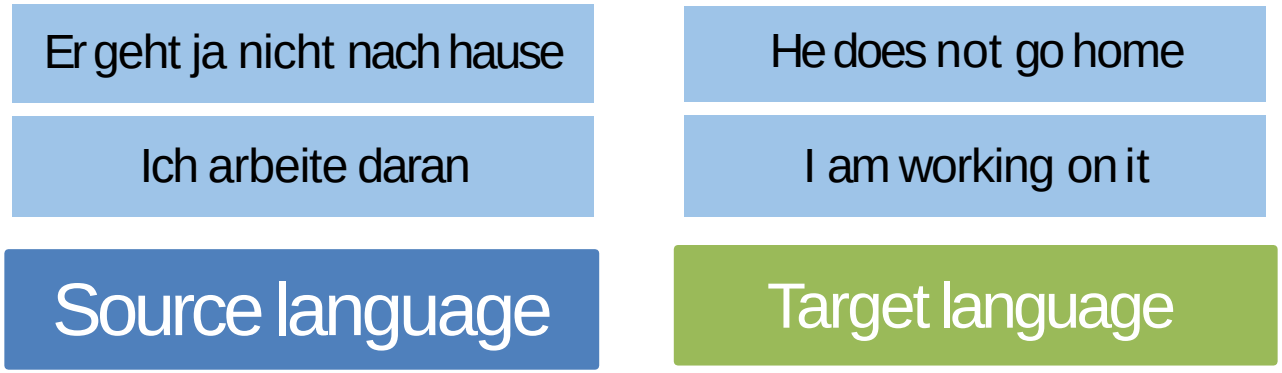
Sequence to Sequence Models

- Machine translation as an example
  - Pairs of sentences in two languages
  - Given a source sequence of words, predict/generate the target sequence of words



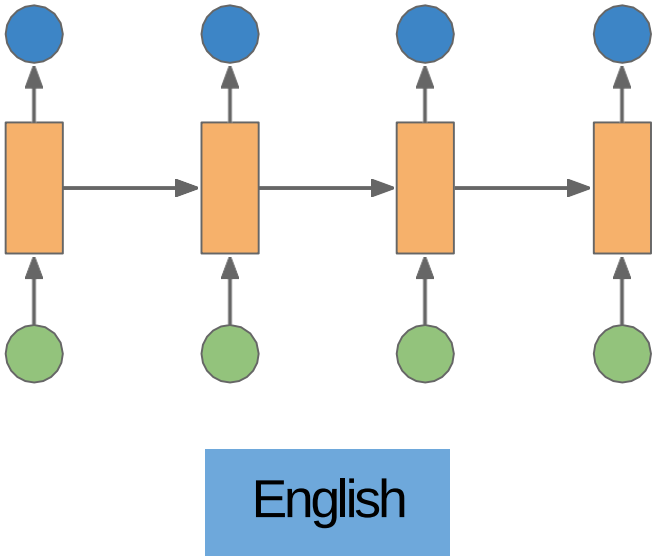
Sequence to Sequence Models

- Machine translation as an example
  - Pairs of sentences in two languages
  - Given a source sequence of words, predict/generate the target sequence of words
  - Target sequence can only be generated *after* processing the entire source sequence



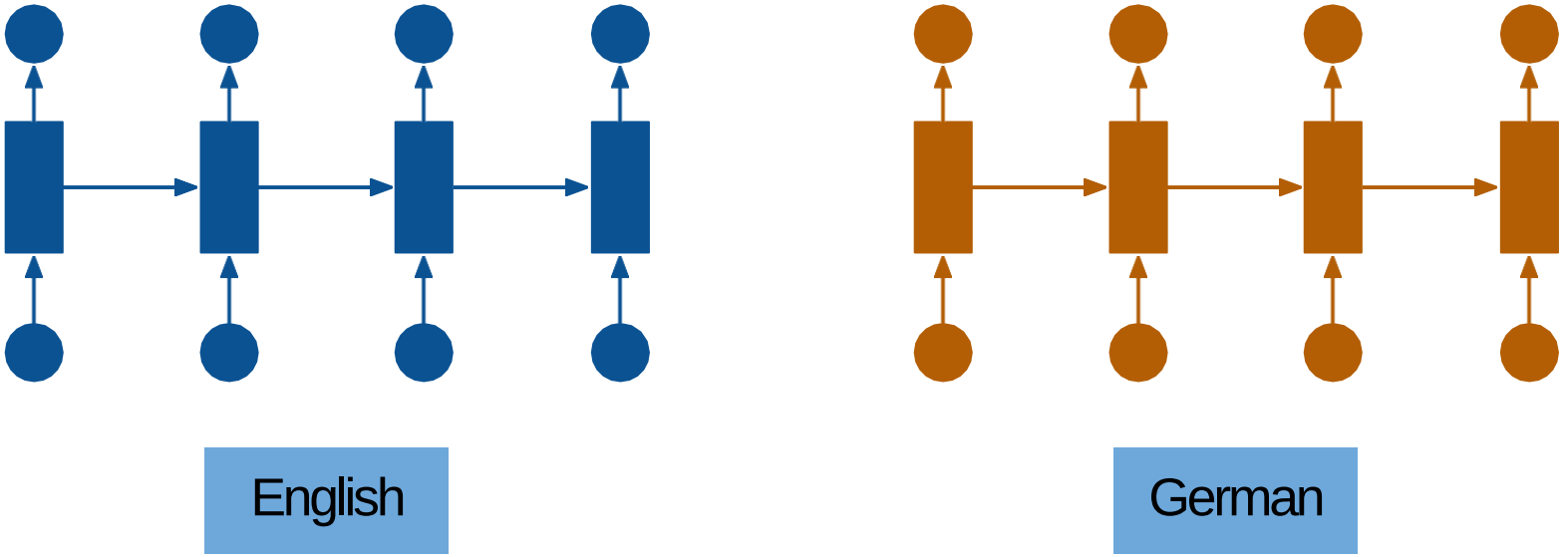
Sequence to Sequence Models

- So far, we have seen a RNN with one language involved:



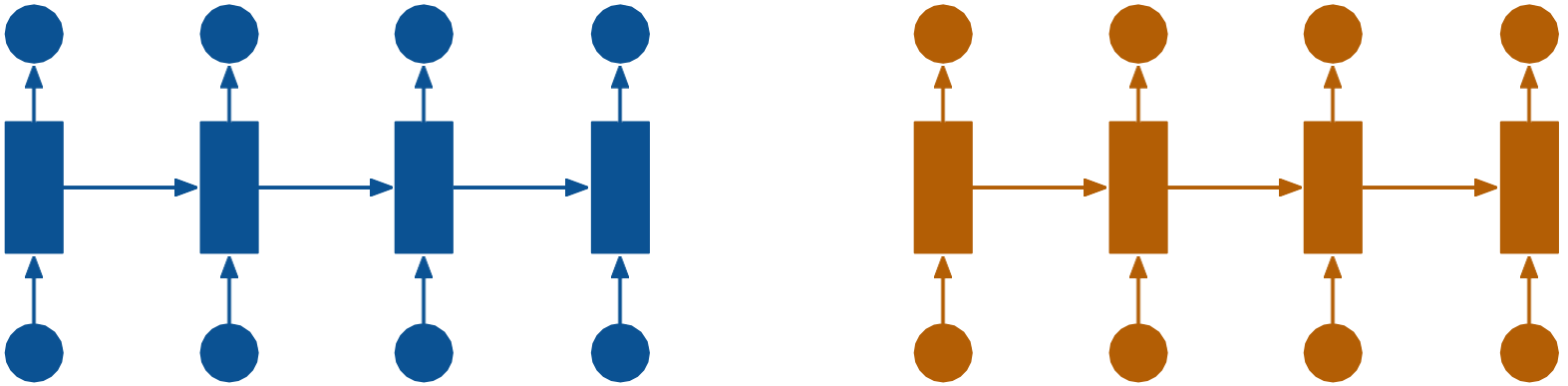
Sequence to Sequence Models

- Simple seq2seq models can be seen as consisting of bilingual RNNs
  - Consider two language models



Sequence to Sequence Models

- **Recap:** a language model looks at a sequence of words and predicts the nextword



Sequence to Sequence Models

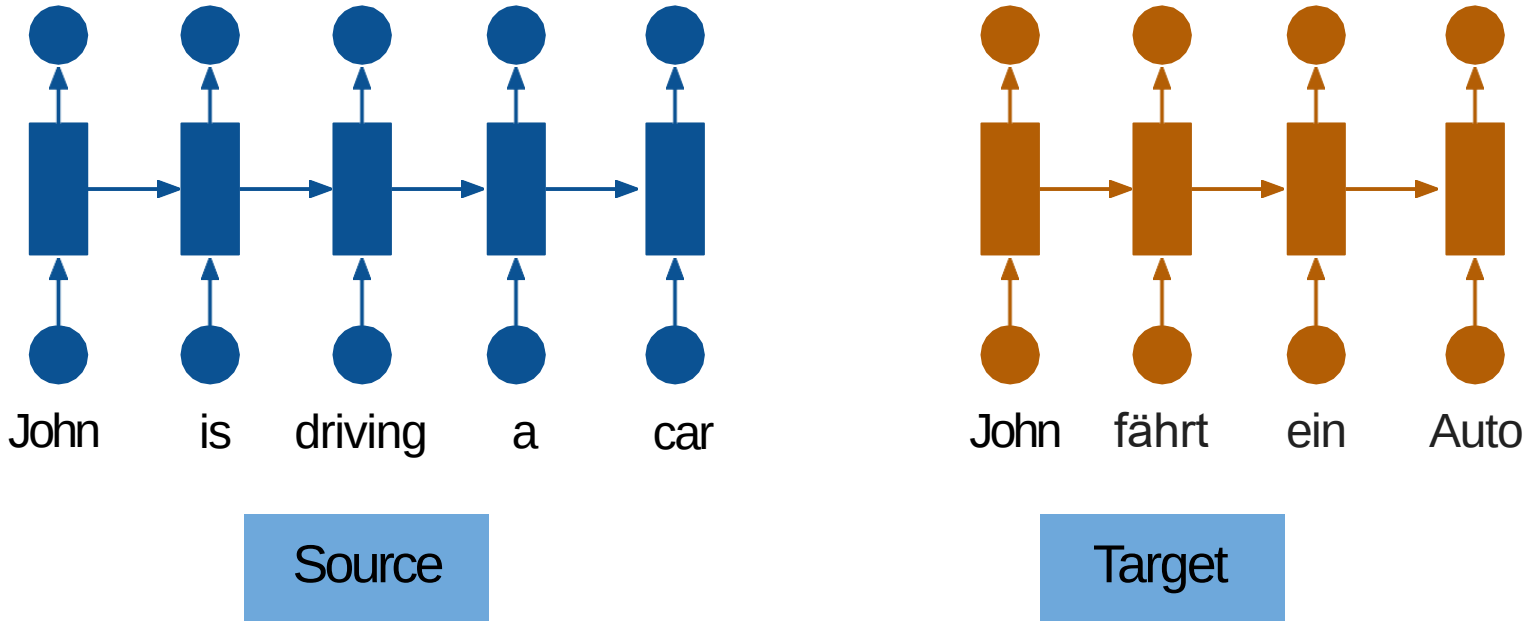
- Now, consider a Bilingual RNN
  - Imagine English and German sentences as a single sequence of strings
  - No explicit information about source and target language

John is driving a car . John fährt ein Auto .

Boundary symbols to mark source and target sentence endings

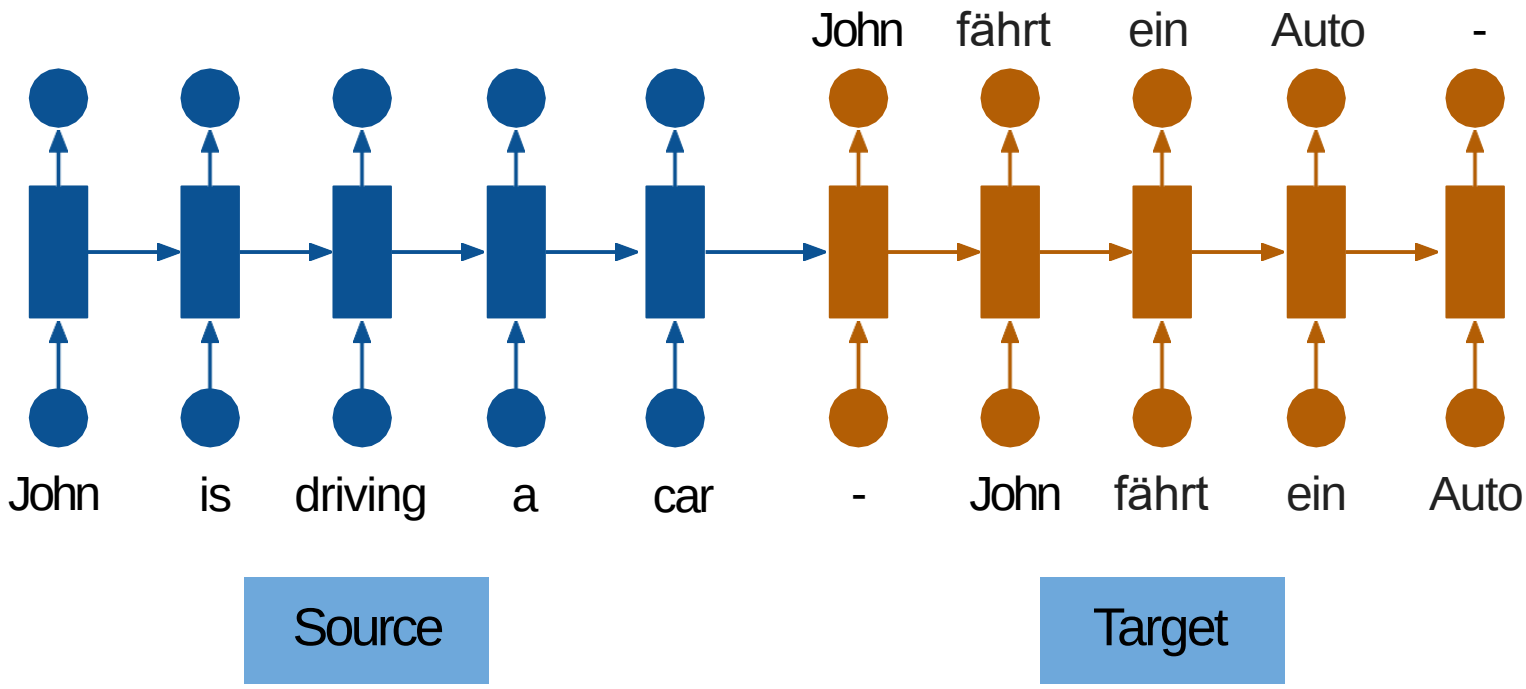
Sequence to Sequence Models

- Consider this combined form as a single language



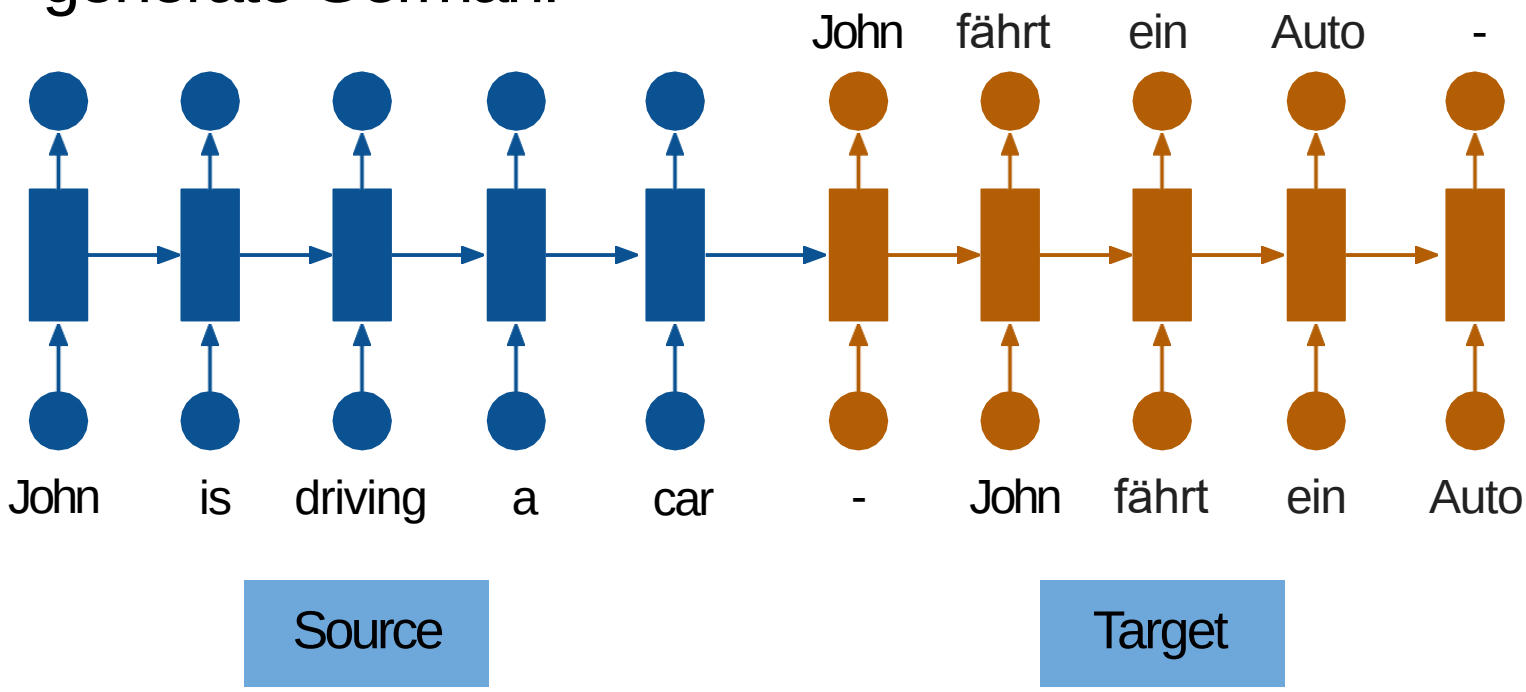
Sequence to Sequence Models

- Consider this combined form as a single language

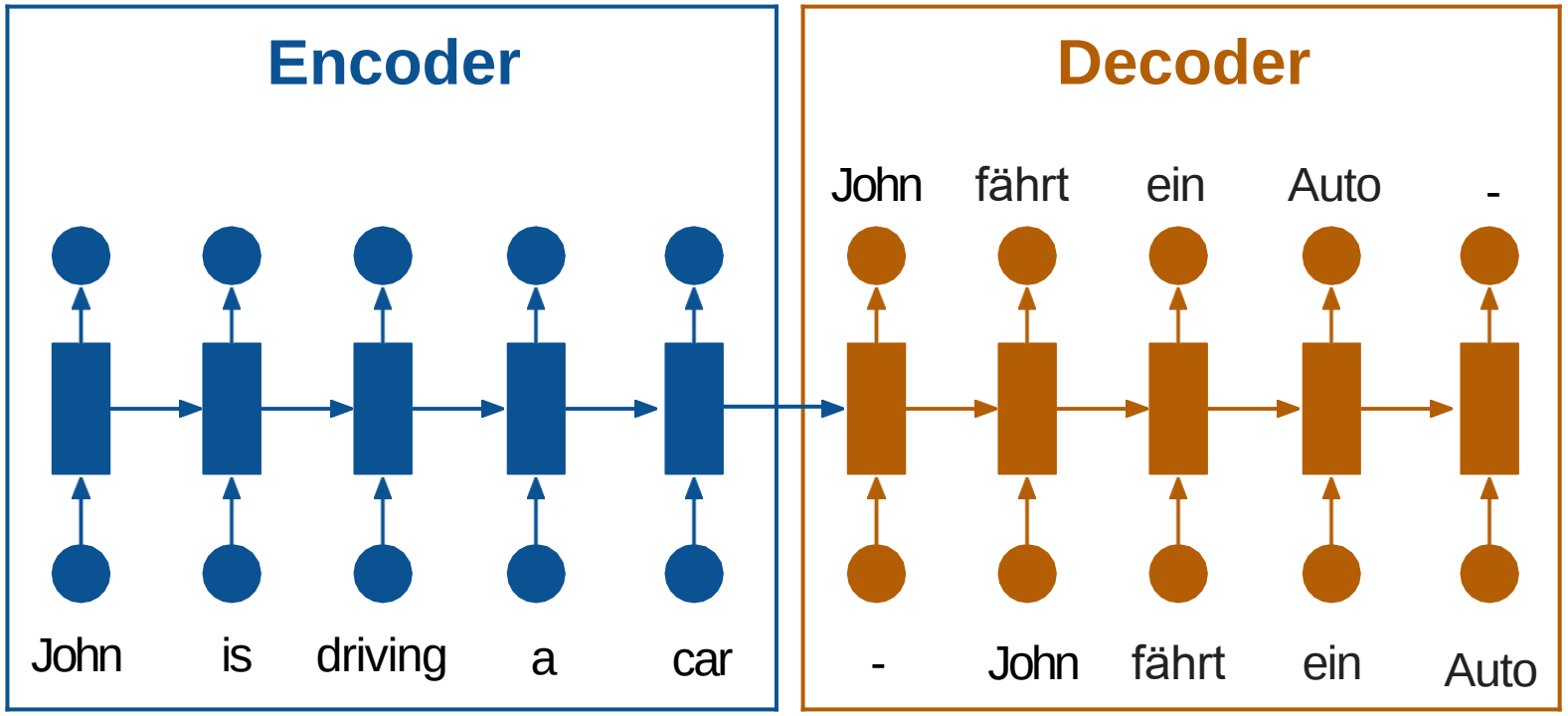


Sequence to Sequence Models

- Essentially, the first RNN is summarizing the English sentence into a vector, and the second RNN uses this to generate German!

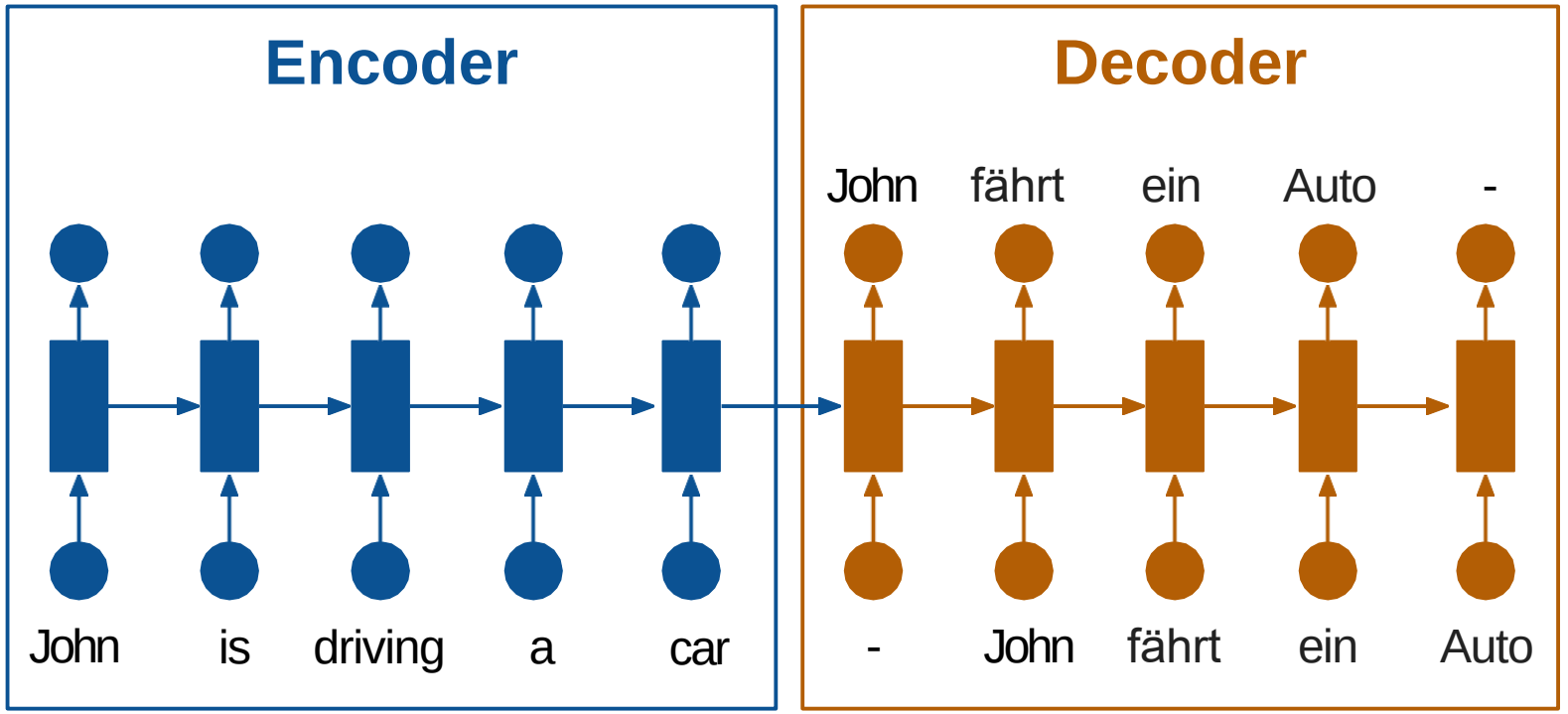


Sequence to Sequence Models



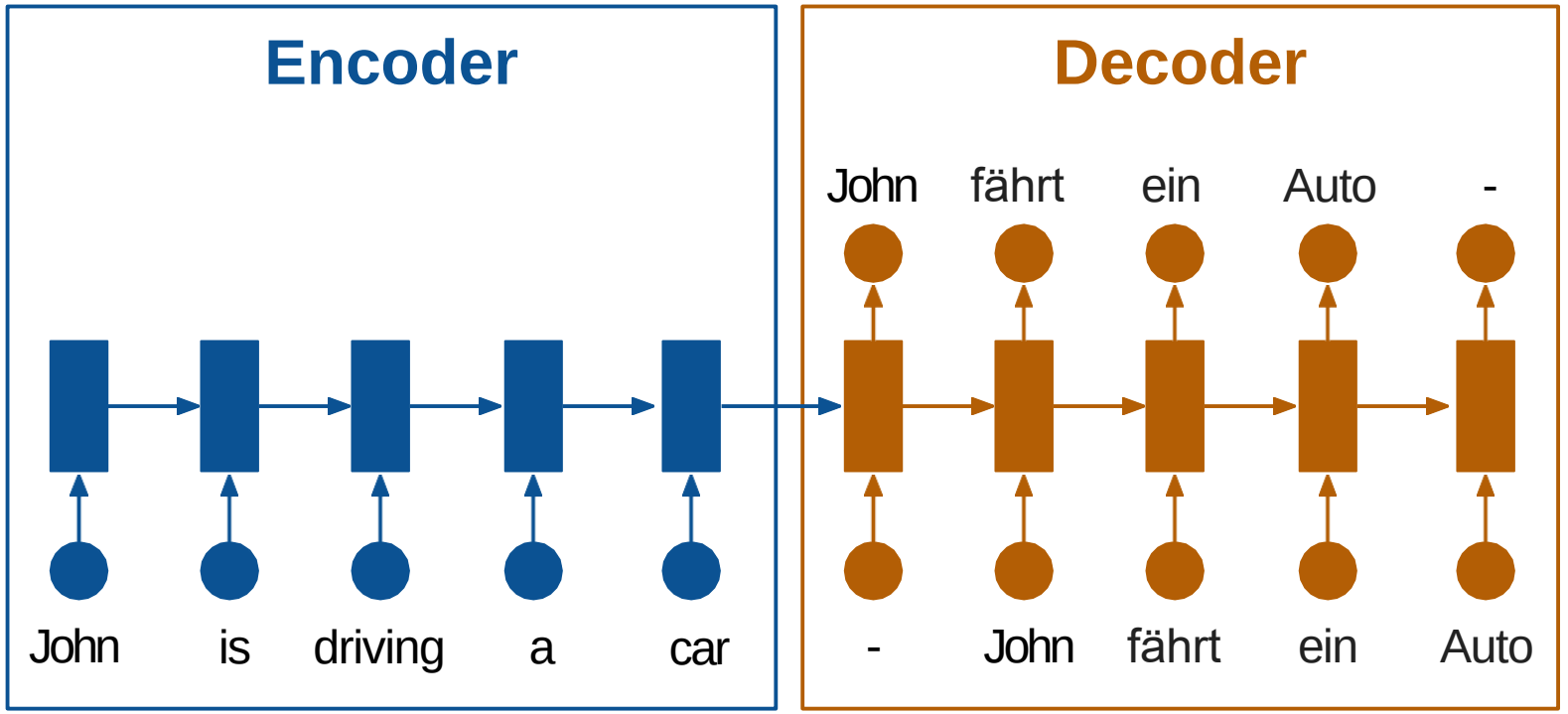
This is also called the Encoder-Decoder architecture!

Sequence to Sequence Models



The model learns to read a source sequence, and then predict the corresponding target sequence

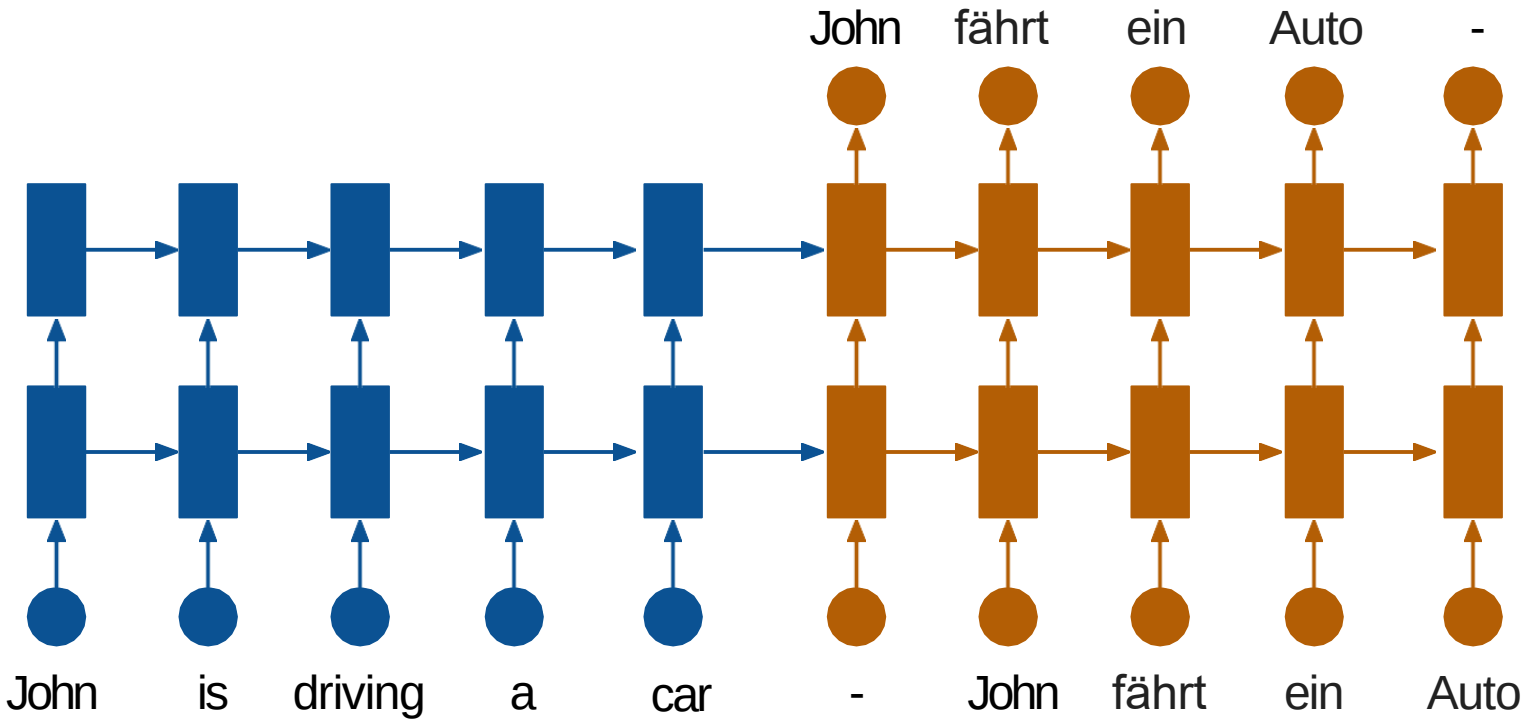
Sequence to Sequence Models



We generally do not produce any outputs in the Encoder, since we are interested in generating the second sequence

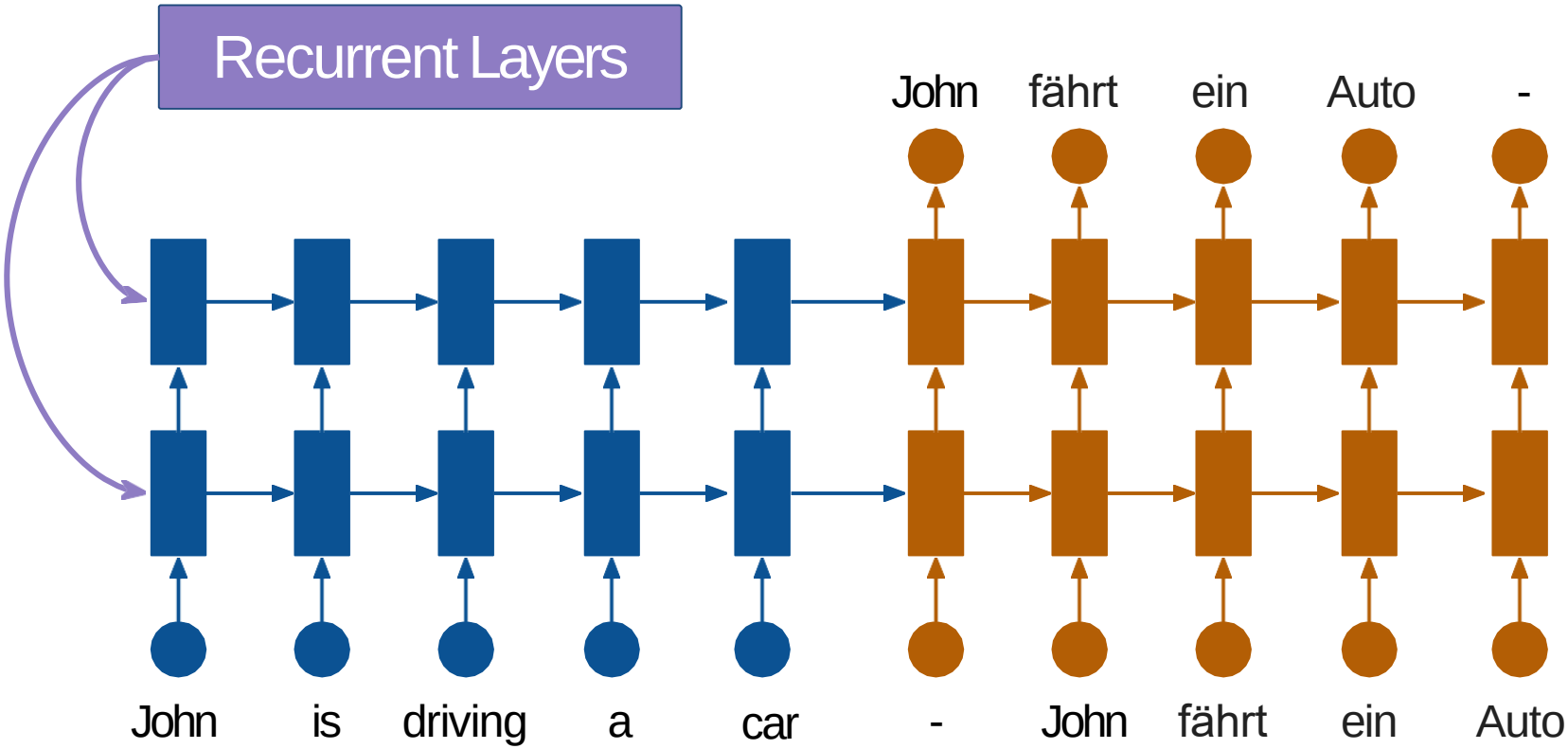
Sequence to Sequence Models

Anatomy of a seq2seq model



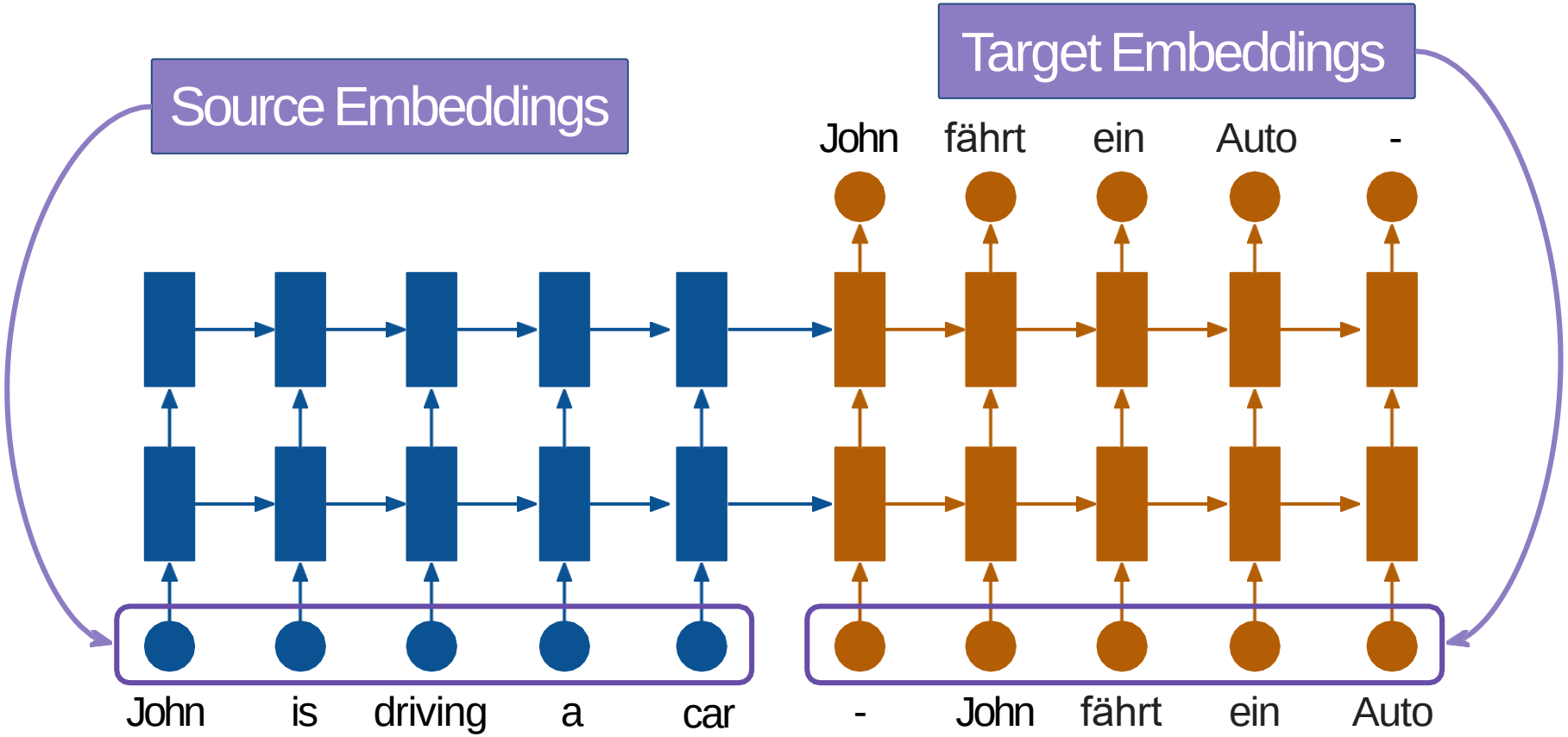
Sequence to Sequence Models

Anatomy of a seq2seq model



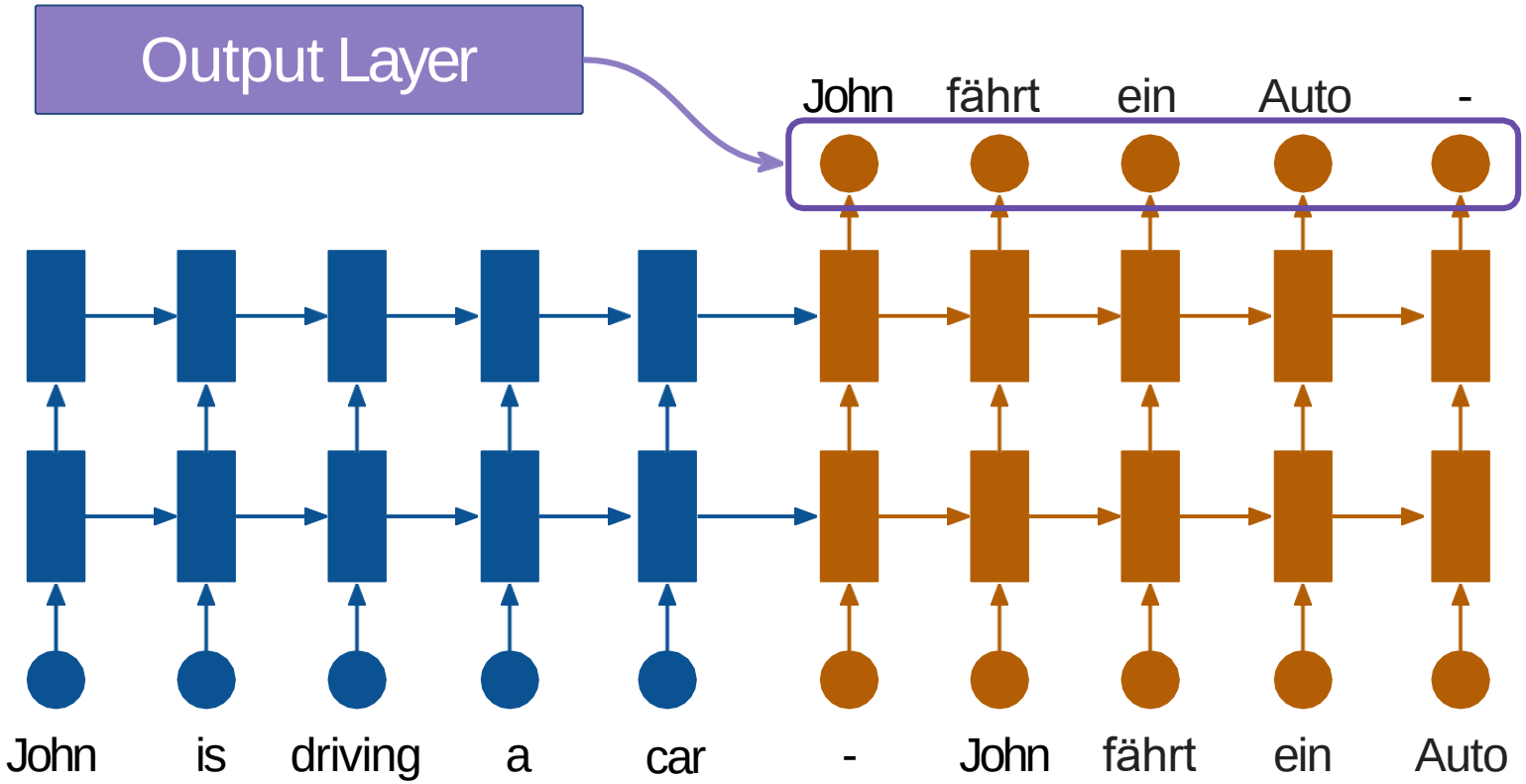
Sequence to Sequence Models

Anatomy of a seq2seq model

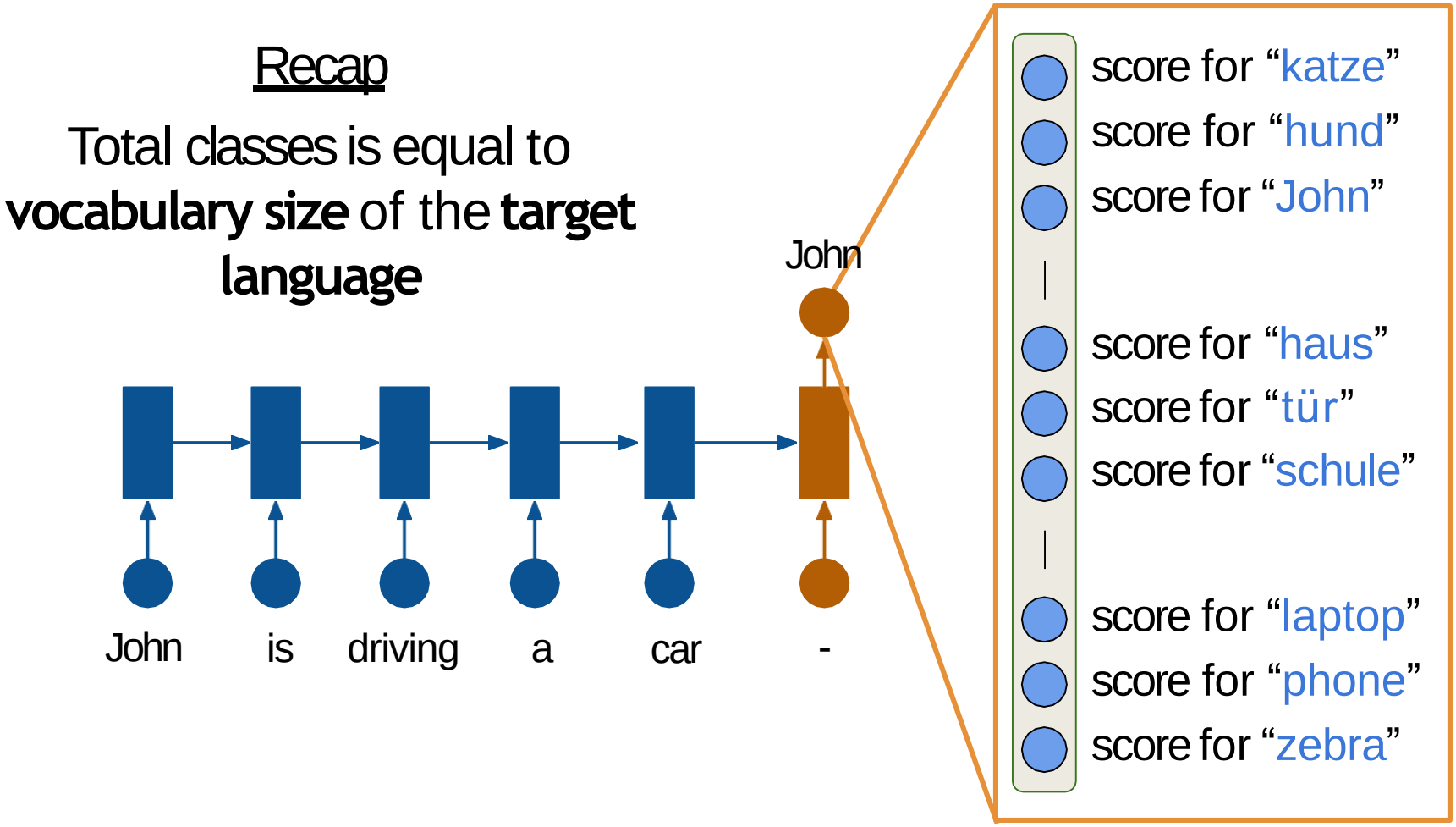


Sequence to Sequence Models

Anatomy of a seq2seq model

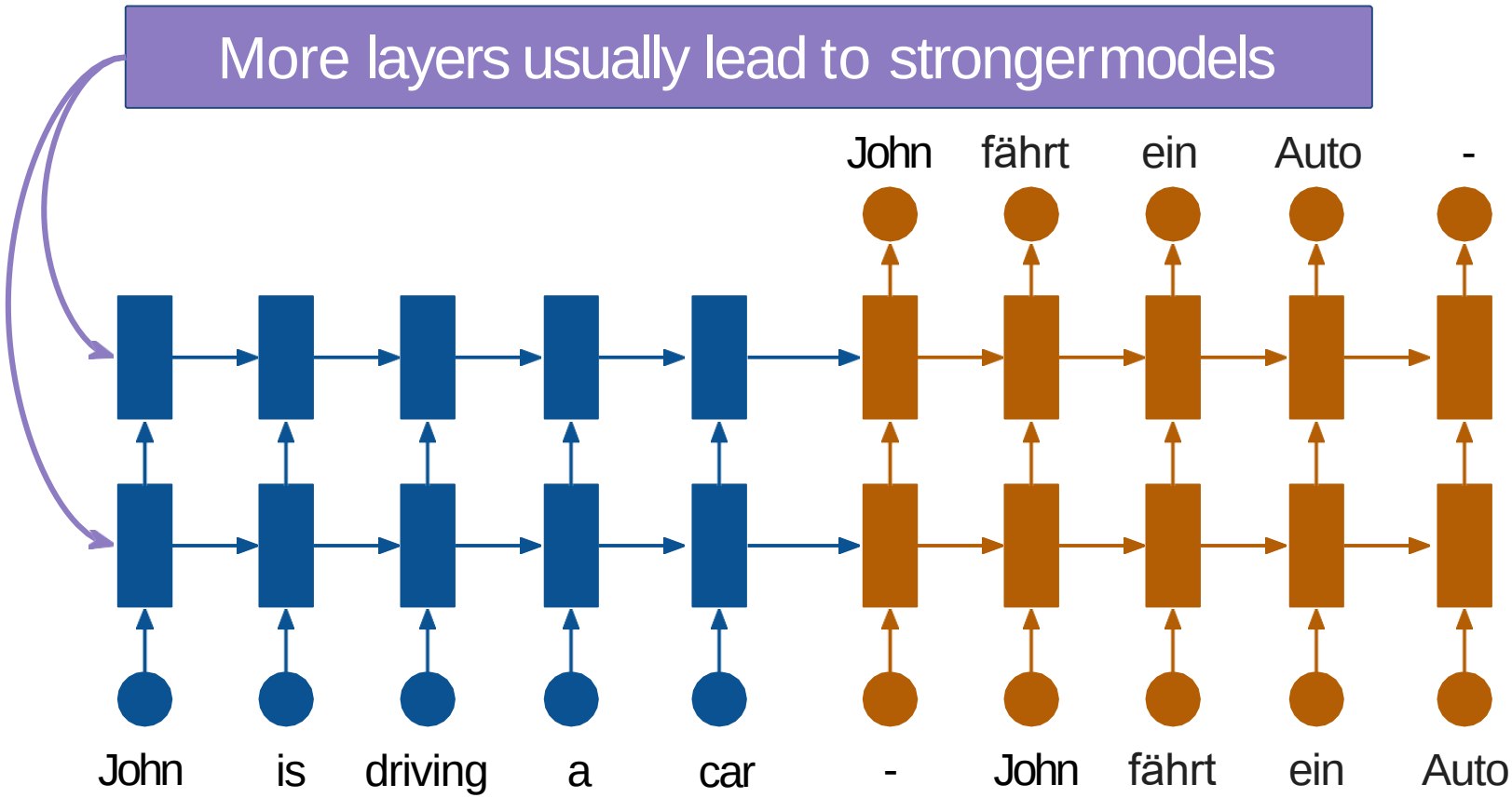


Output Layer



Sequence to Sequence Models

Anatomy of a seq2seq model



Sequence to Sequence Models

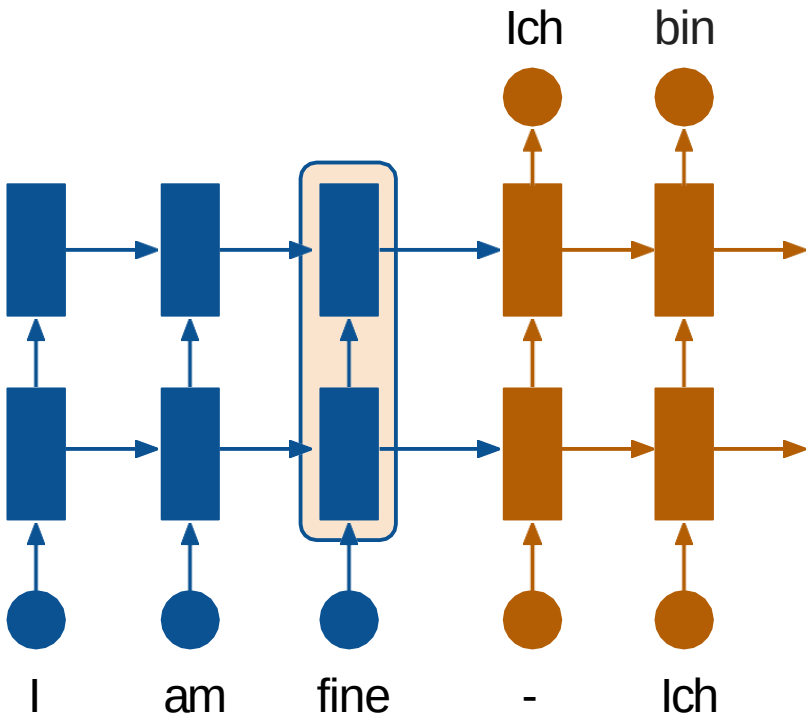
- **End-to-end training:** one loss to optimize all parameters
- **Better use of context:** use source sentences and previously predicted target words
- **Distributed word representations**: semantic similarities

Problems with Sequence to Sequence Models

Problems with Seq2Seq Models

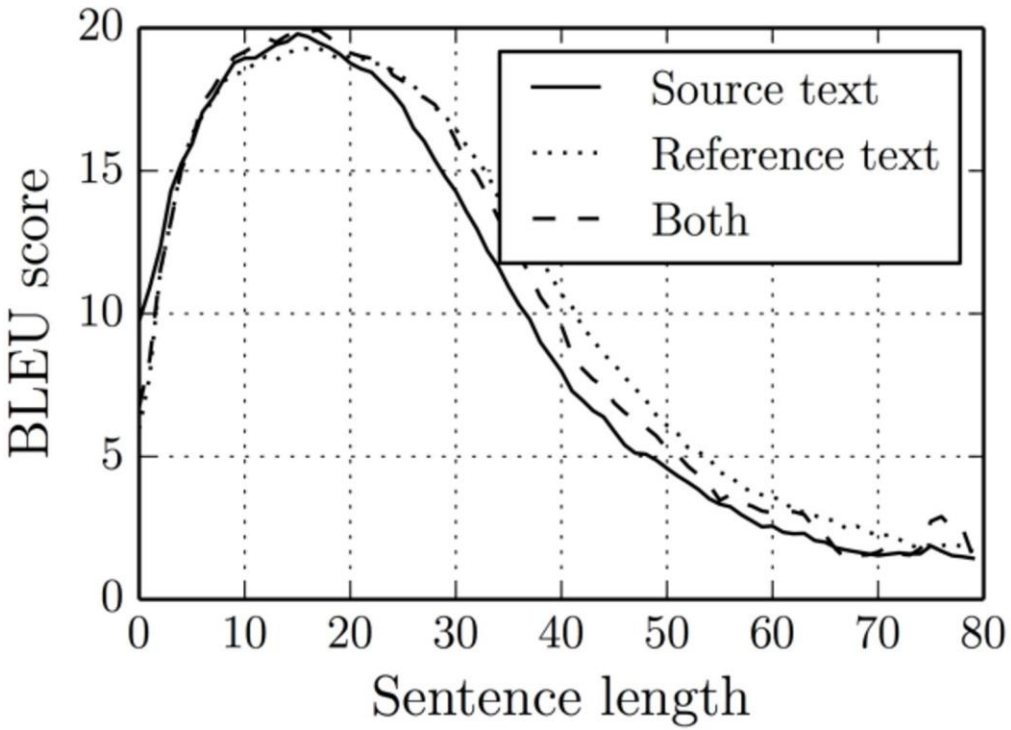
- Sequence to sequence models perform poorly when translating long sentences

- **Issue**: a sentence is represented as a fixed vector
- Relationship between source and target words is very abstract
- All words are not equally important to predict a target word



Problems with Seq2Seq Models

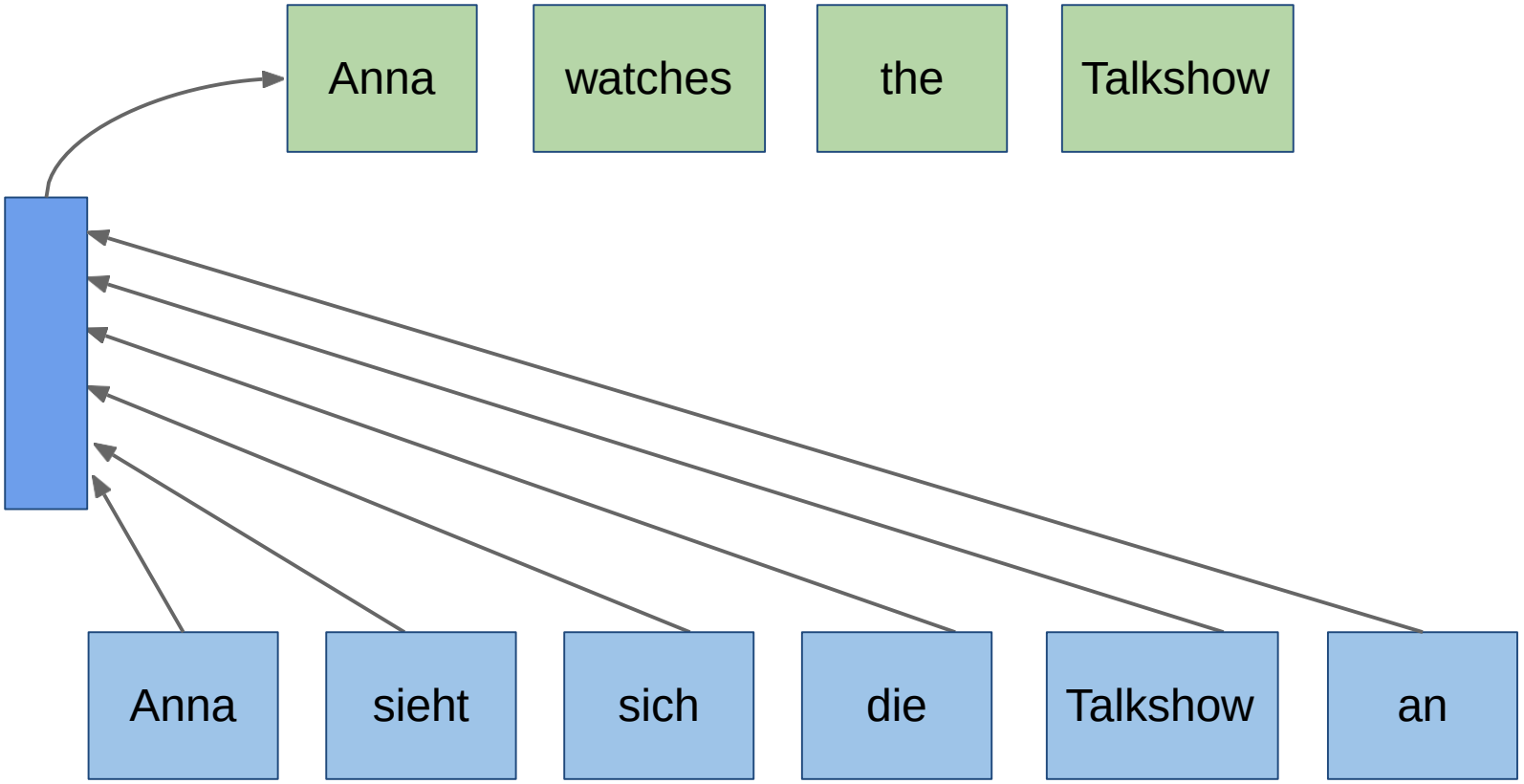
- Sequence to sequence models perform poorly when translating long sentences



https://devblogs.nvidia.com/parallelforall/introduction-neural-machine-translation-gpus-part-3/

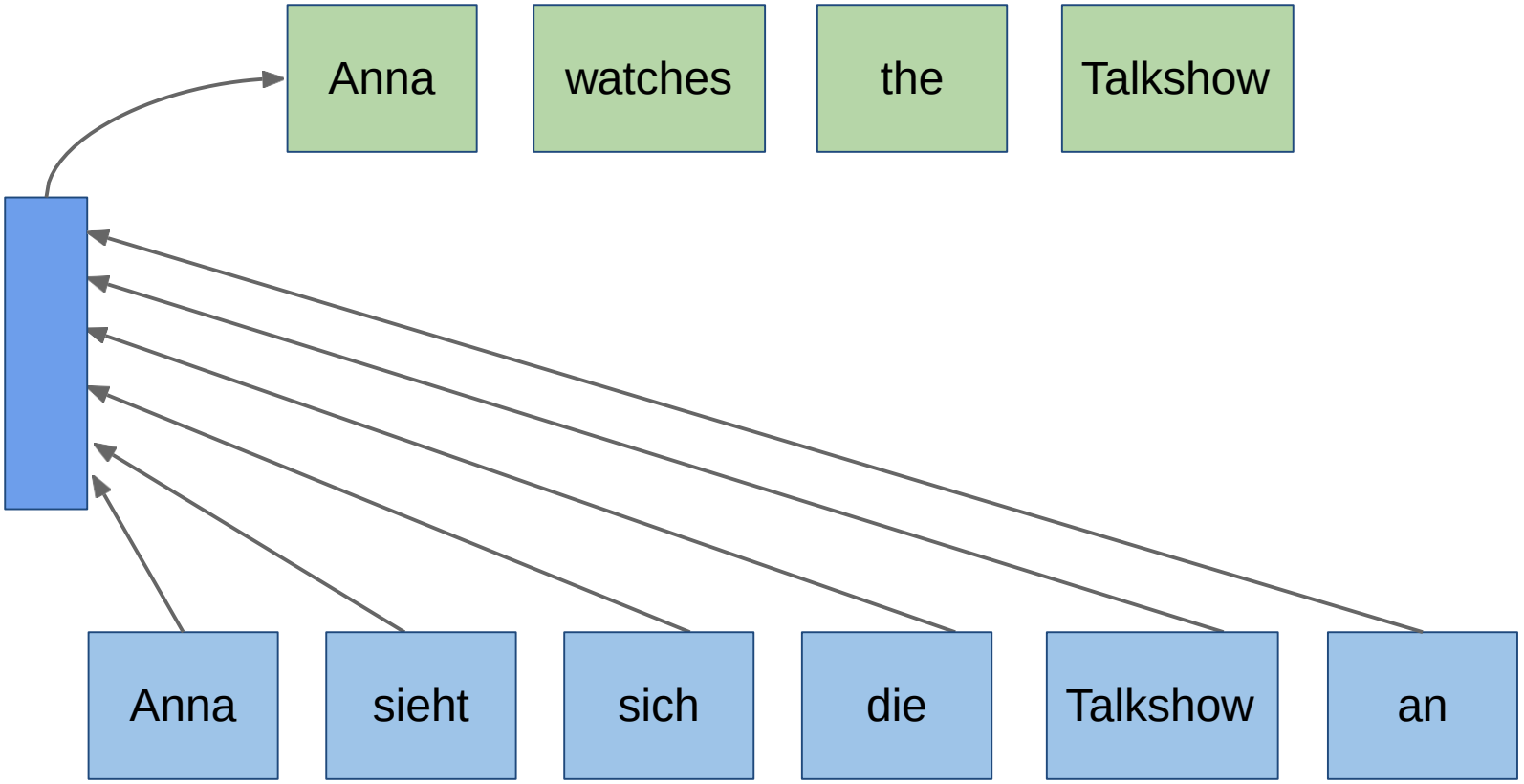
Attention Mechanism

One vector represents all source words



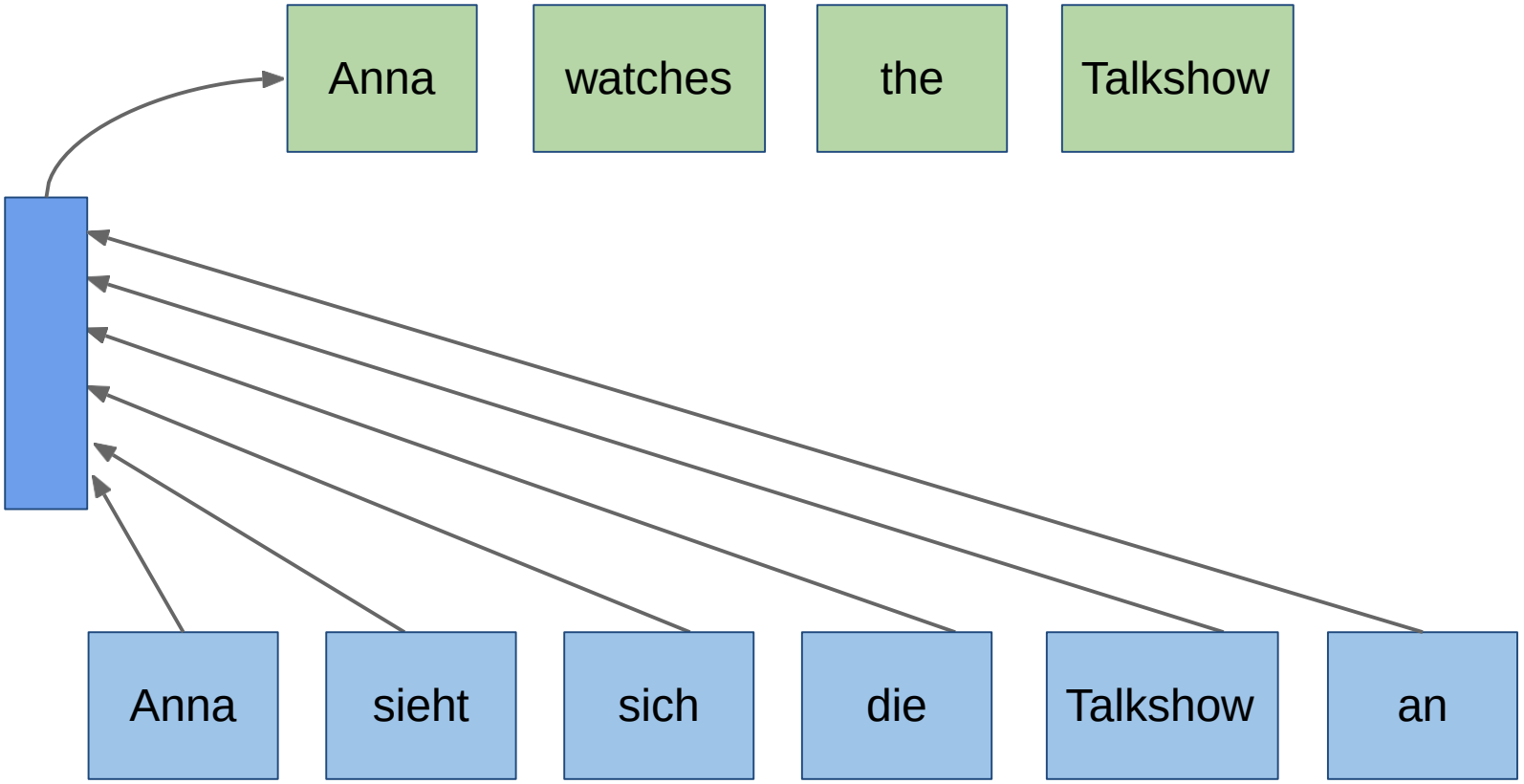
Attention Mechanism

Source and target words inherently have relationships



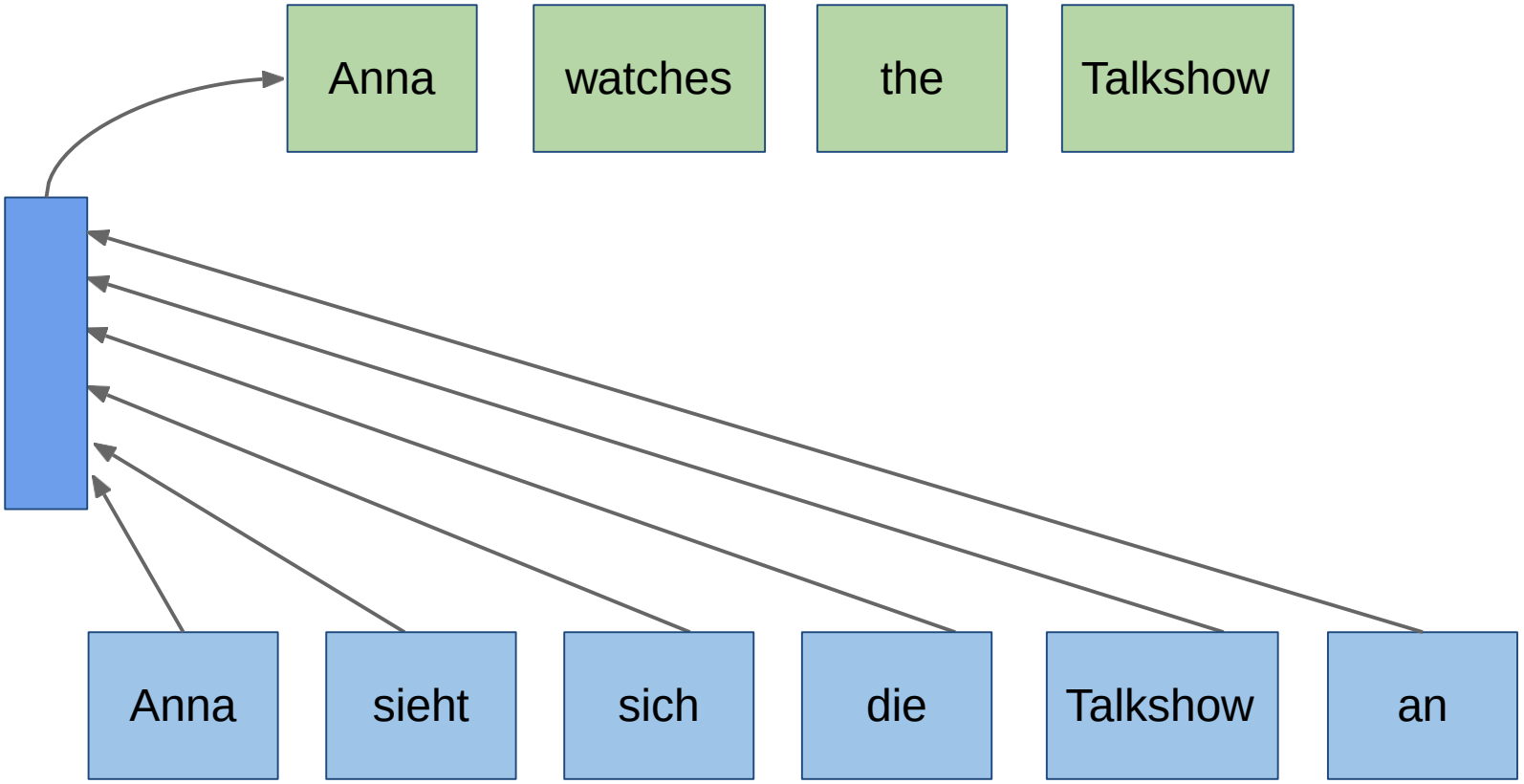
Attention Mechanism

Anna,Anna are more relevant to each other than Anna,Talkshow



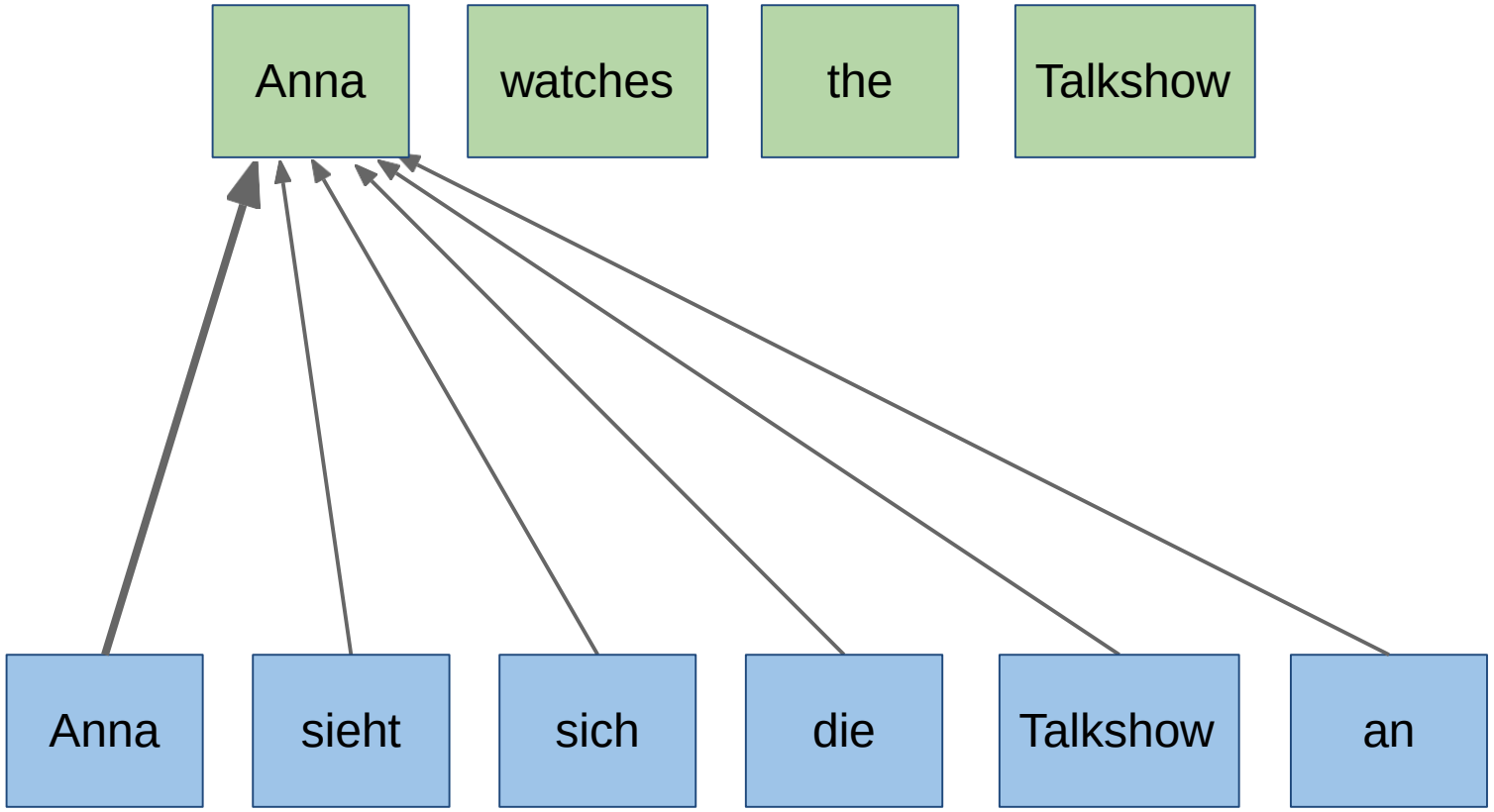
Attention Mechanism

Solution: For each target word, concentrate only on specific source words



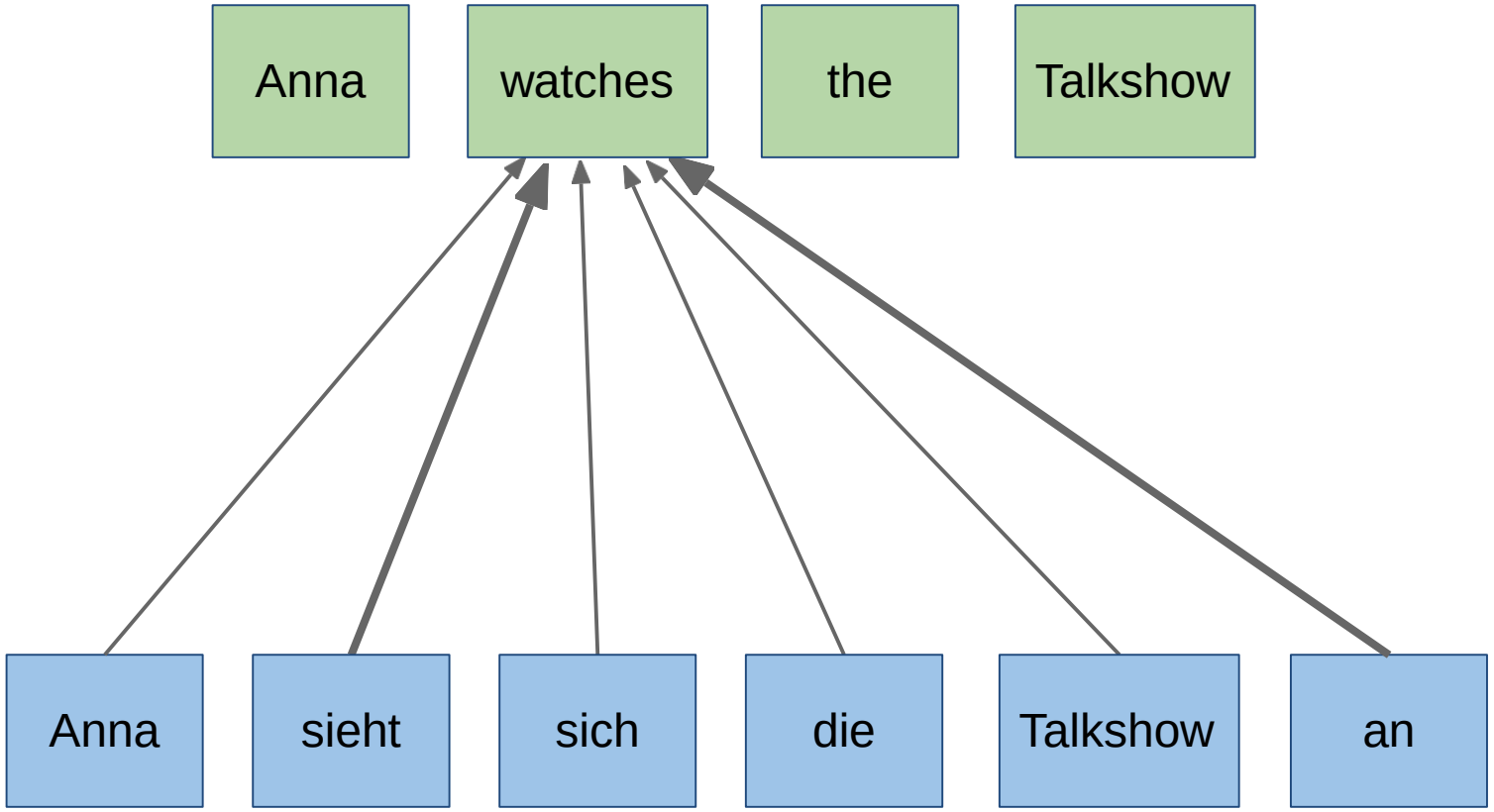
Attention Mechanism

Solution: For each target word, concentrate only on specific source words



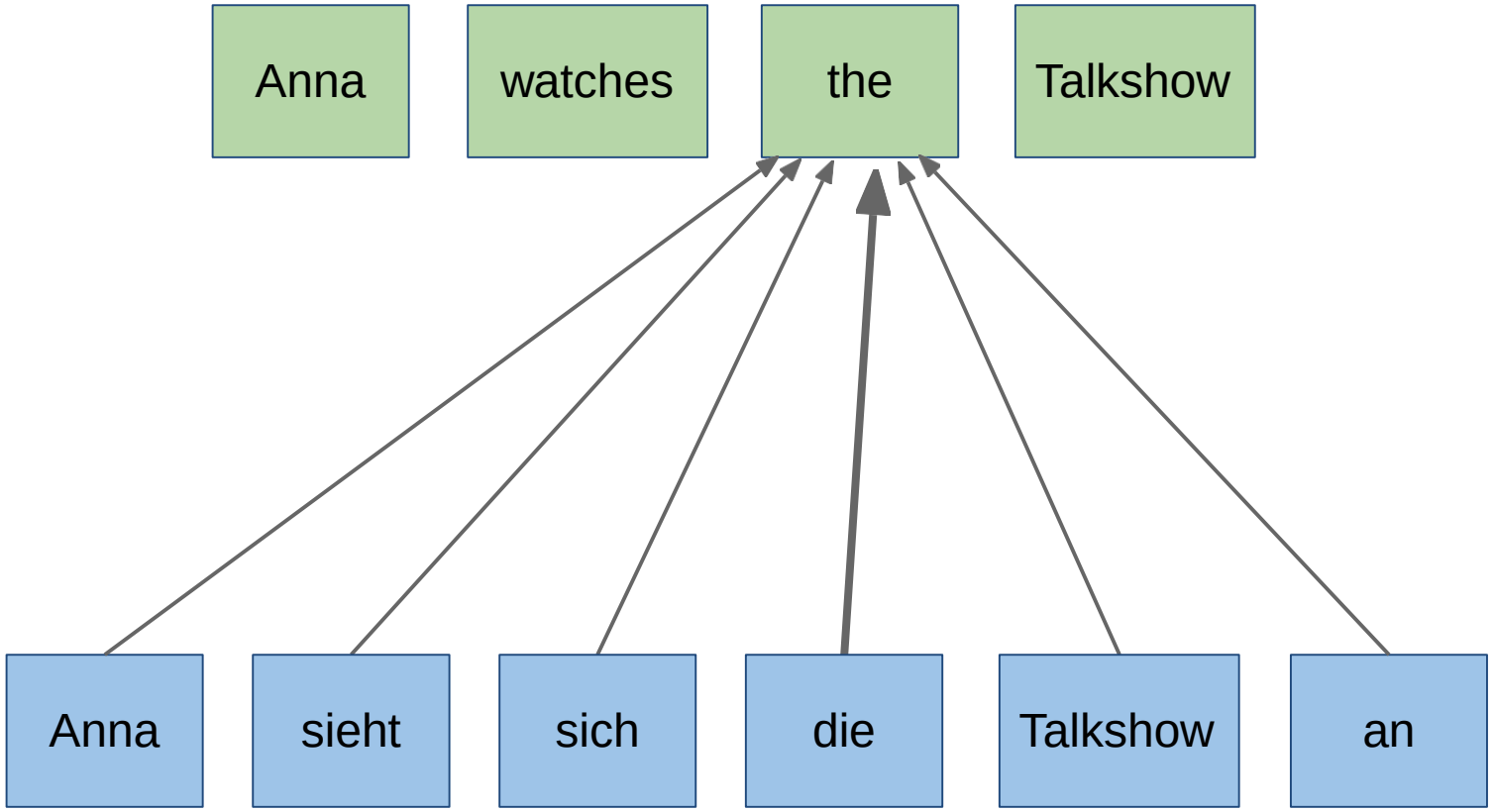
Attention Mechanism

Solution: For each target word, concentrate only on specific source words



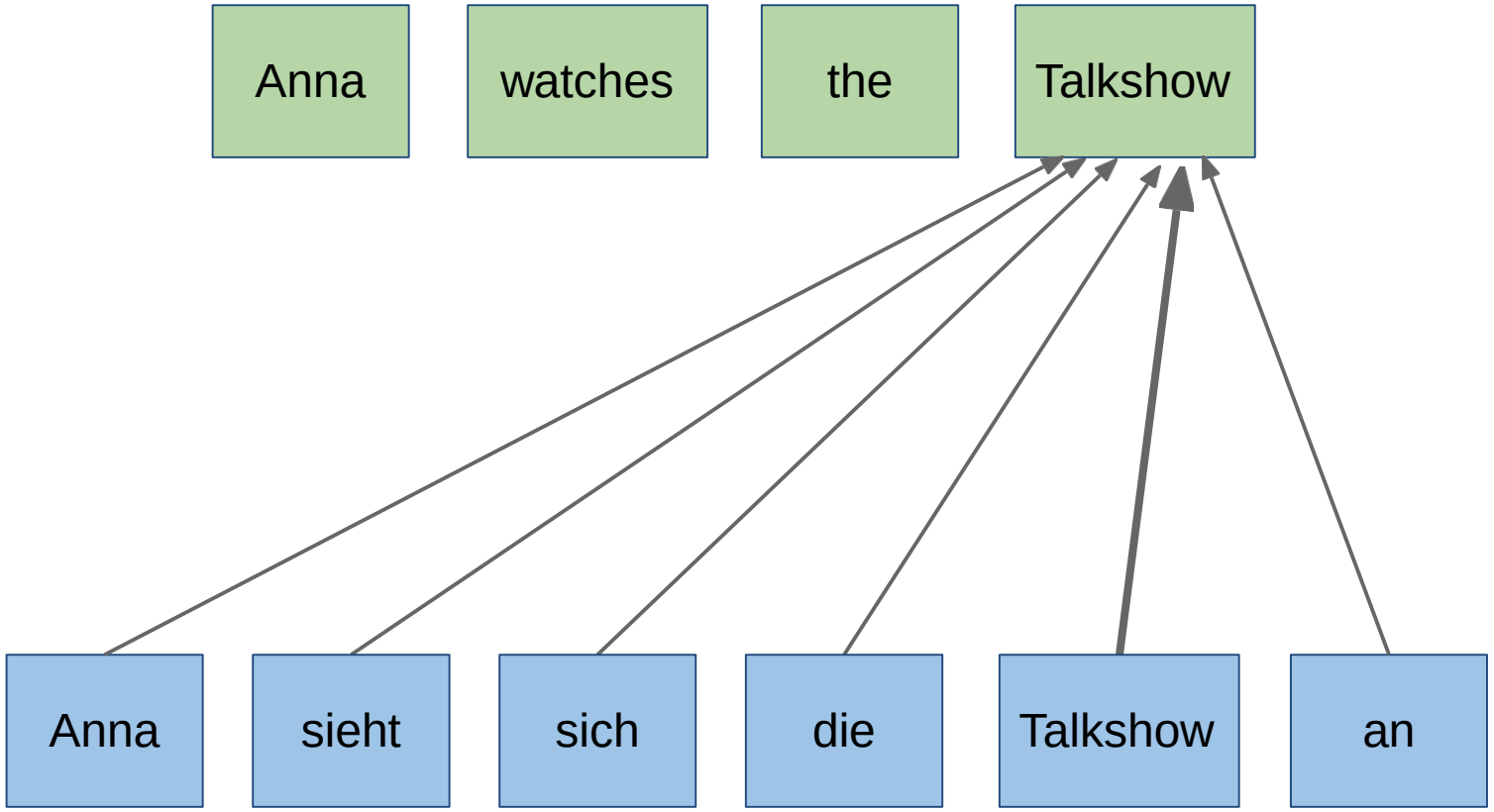
Attention Mechanism

Solution: For each target word, concentrate only on specific source words



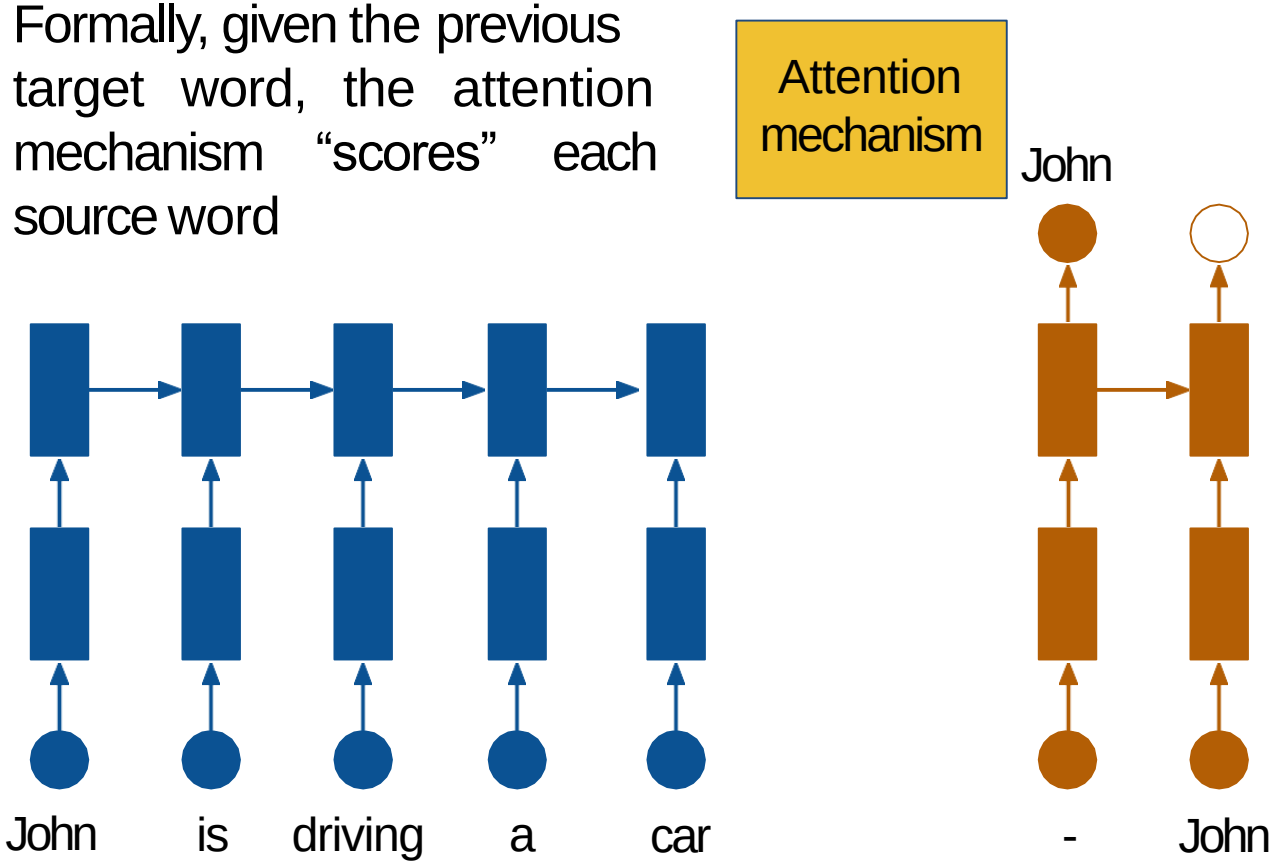
Attention Mechanism

Solution: For each target word, concentrate only on specific source words



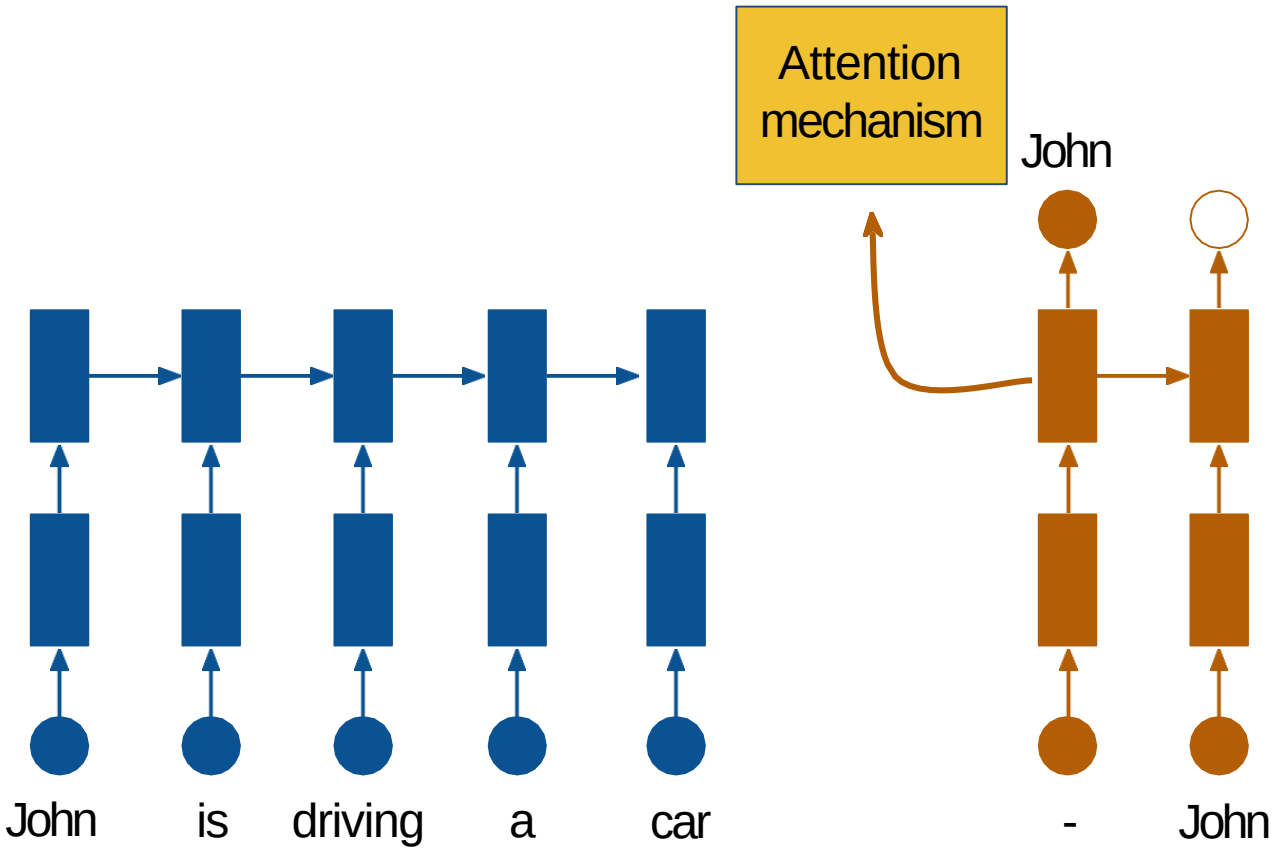
Attention Mechanism

Formally, given the previous target word, the attention mechanism “scores” each source word



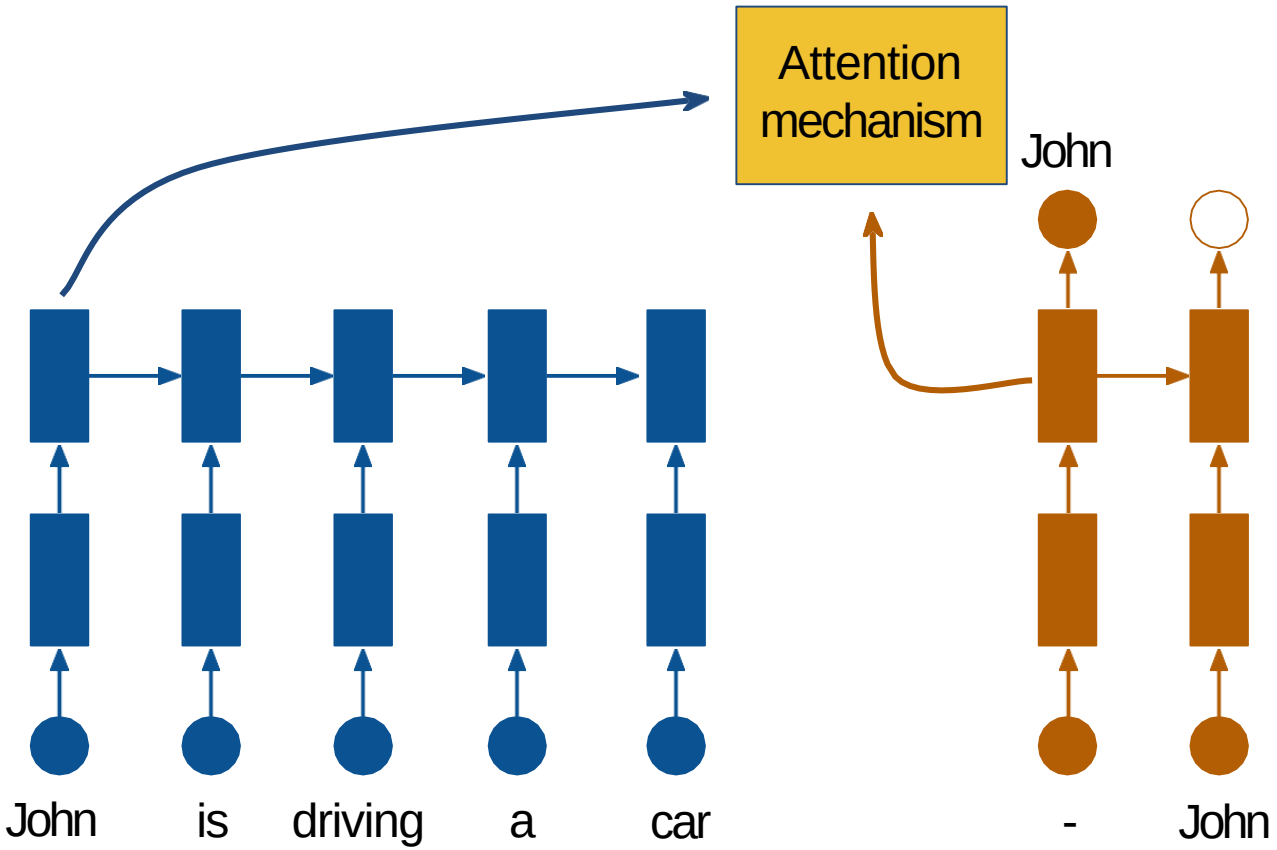
Neural Machine Translation by Luong, Cho and Manning

Attention Mechanism



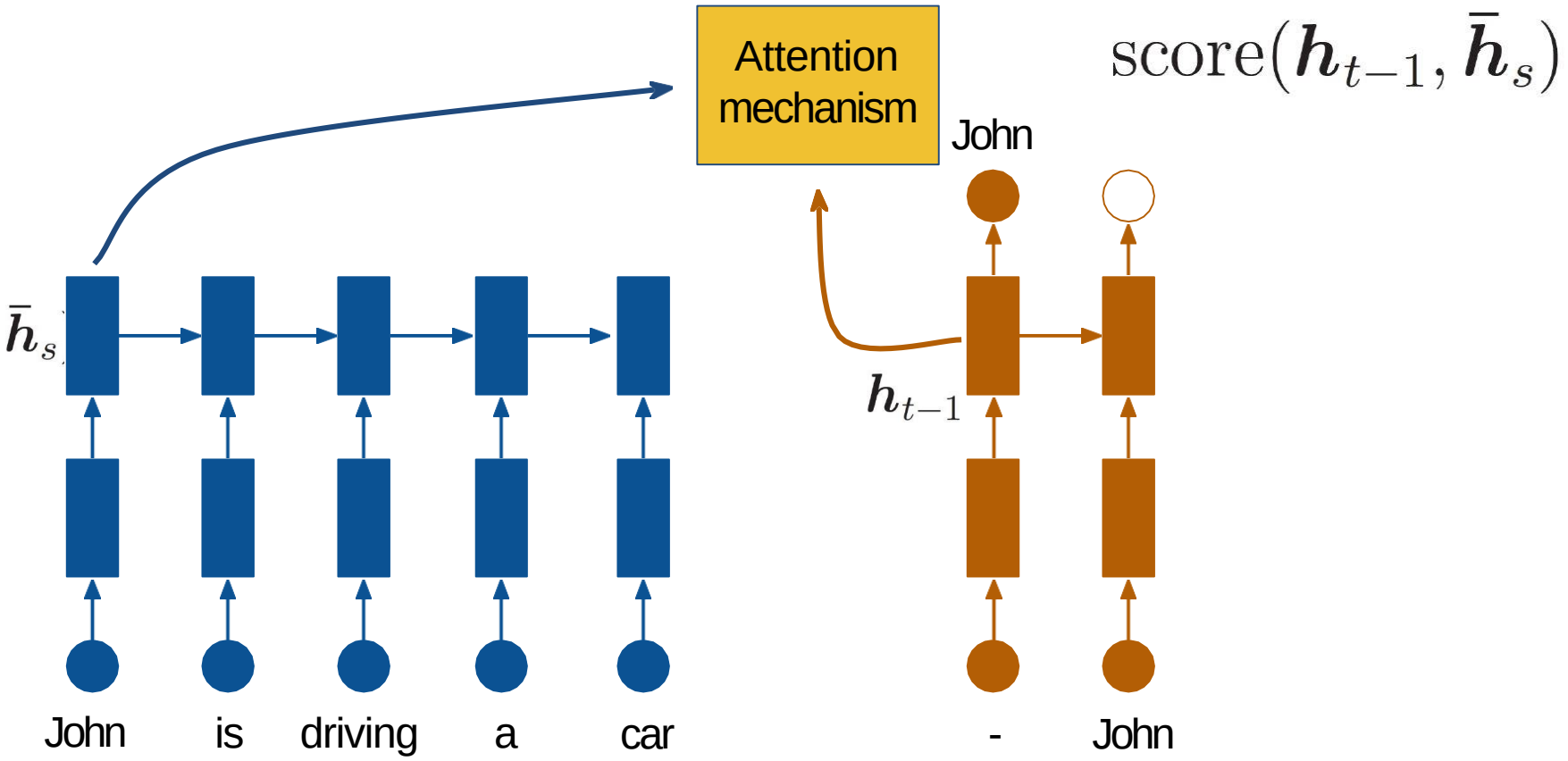
Neural Machine Translation by Luong, Cho and Manning

Attention Mechanism



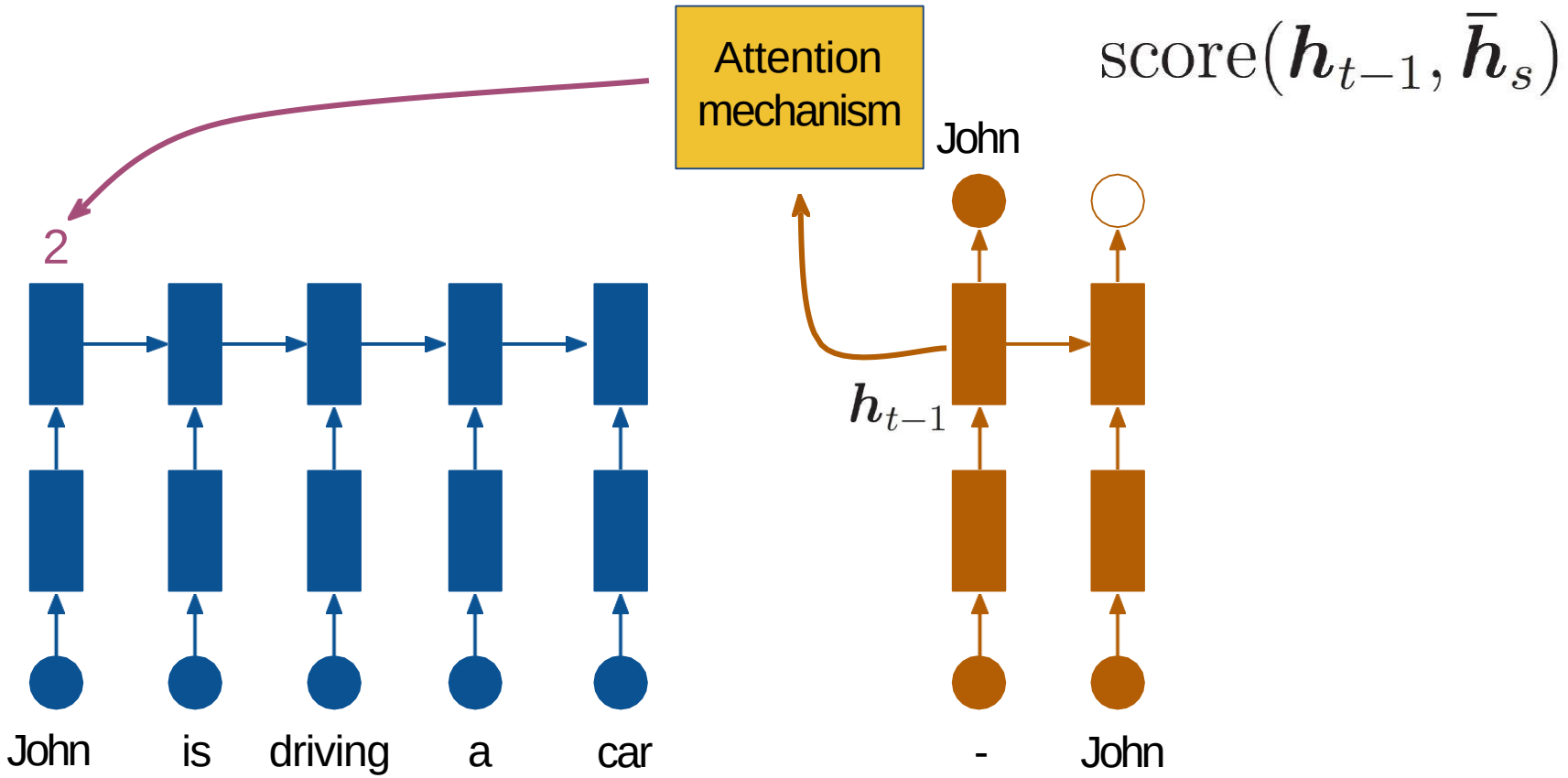
Neural Machine Translation by Luong, Cho and Manning

Attention Mechanism



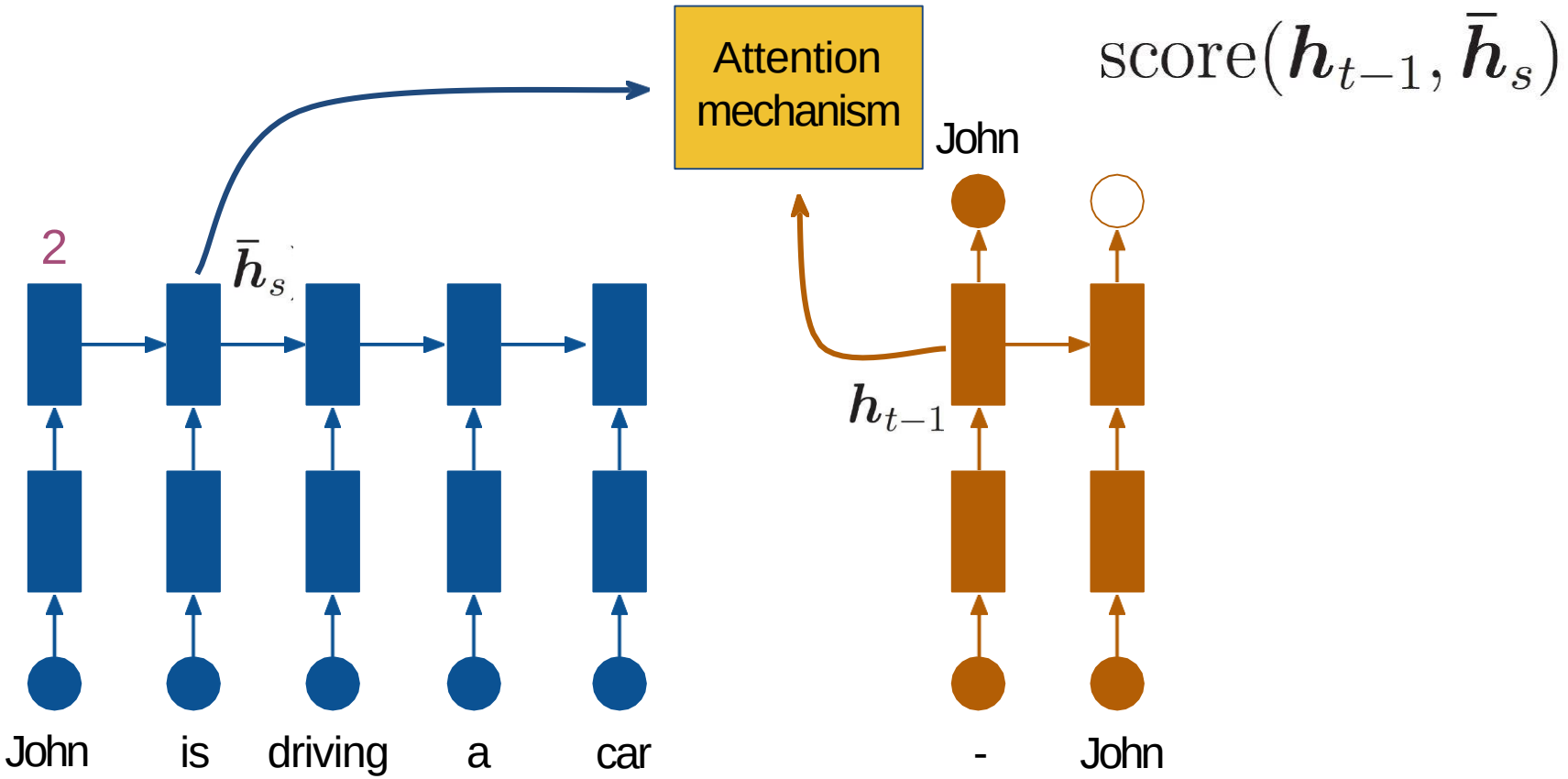
Neural Machine Translation by Luong, Cho and Manning

Attention Mechanism



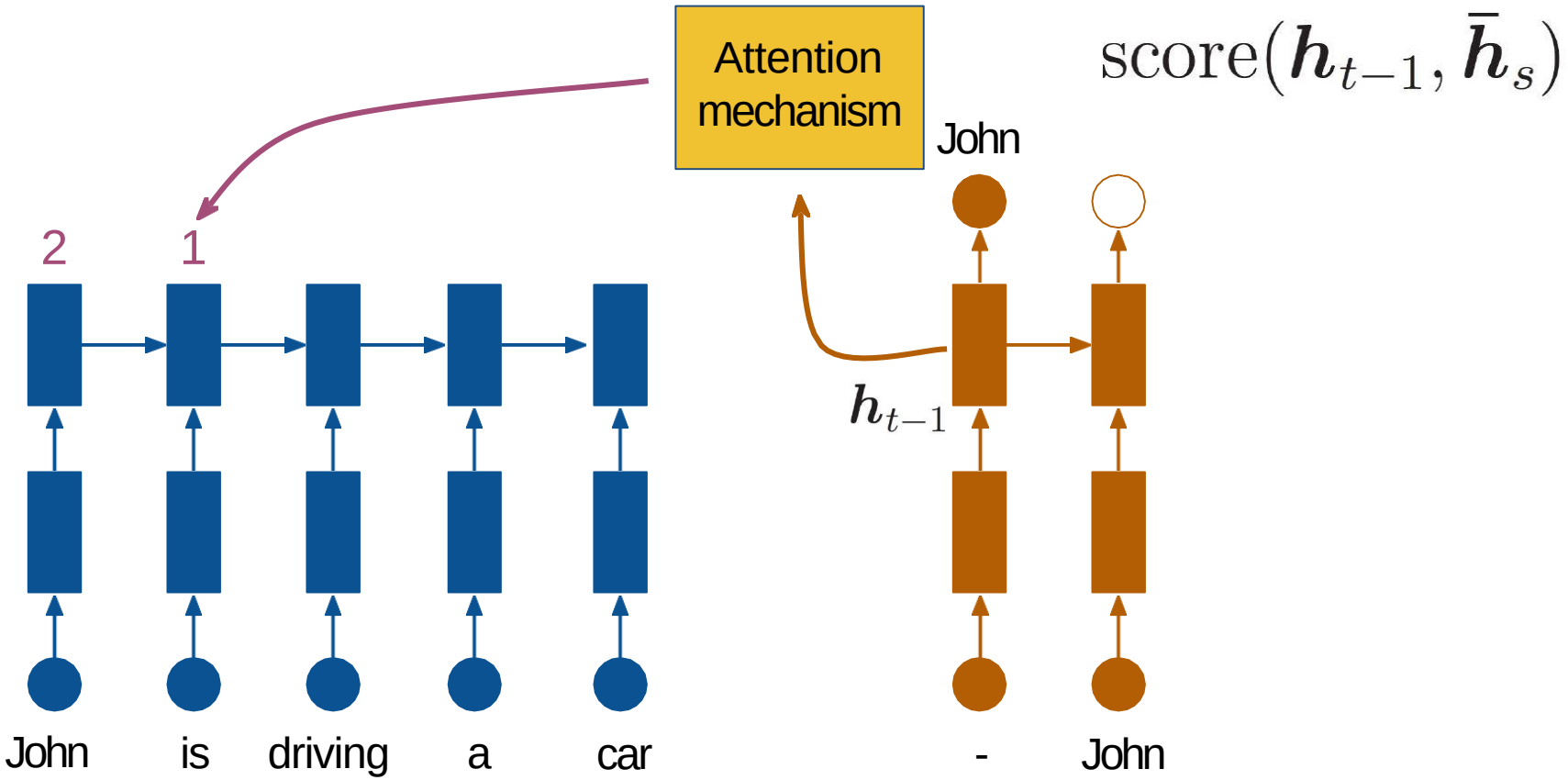
Neural Machine Translation by Luong, Cho and Manning

Attention Mechanism



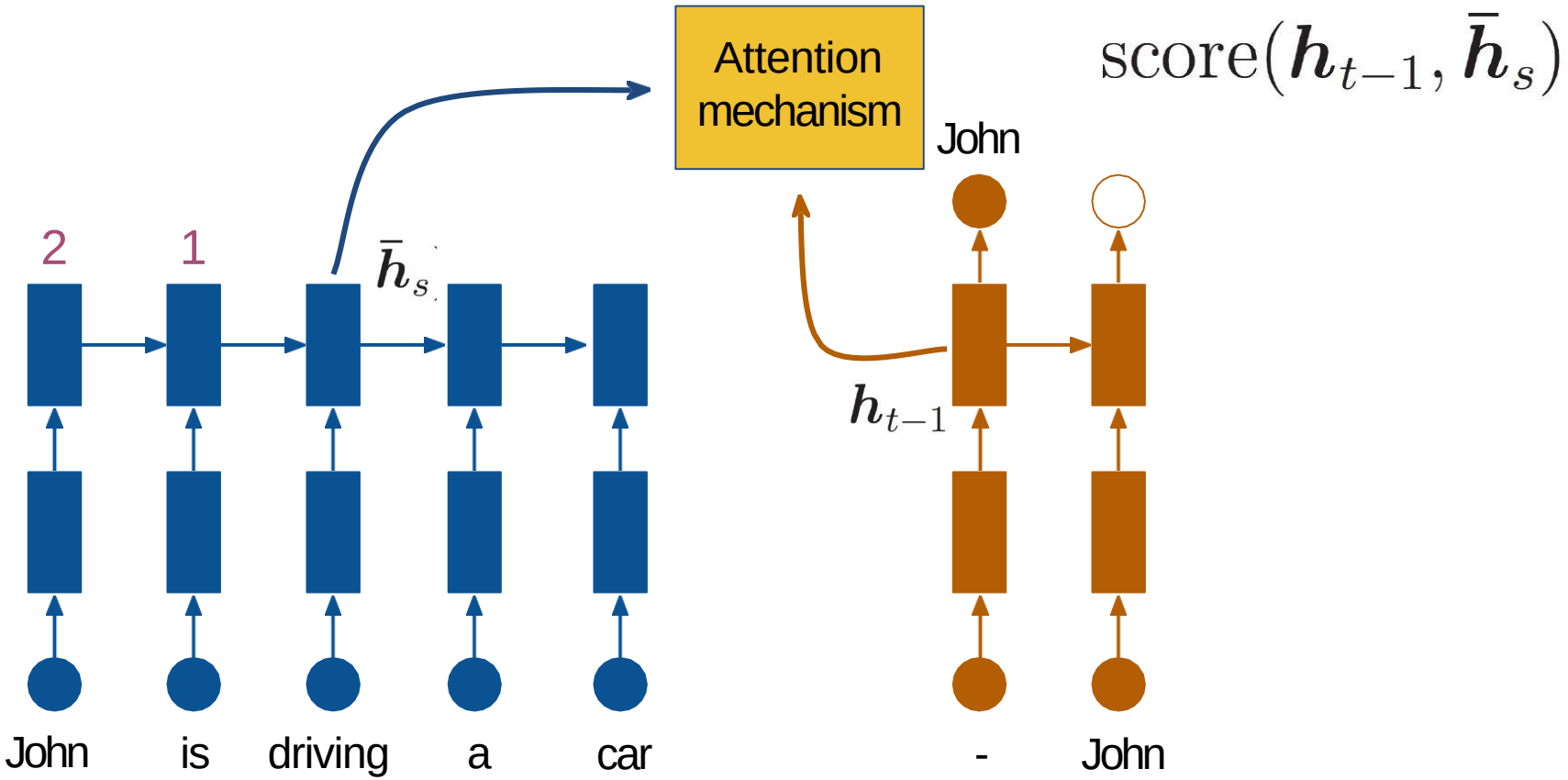
Neural Machine Translation by Luong, Cho and Manning

Attention Mechanism



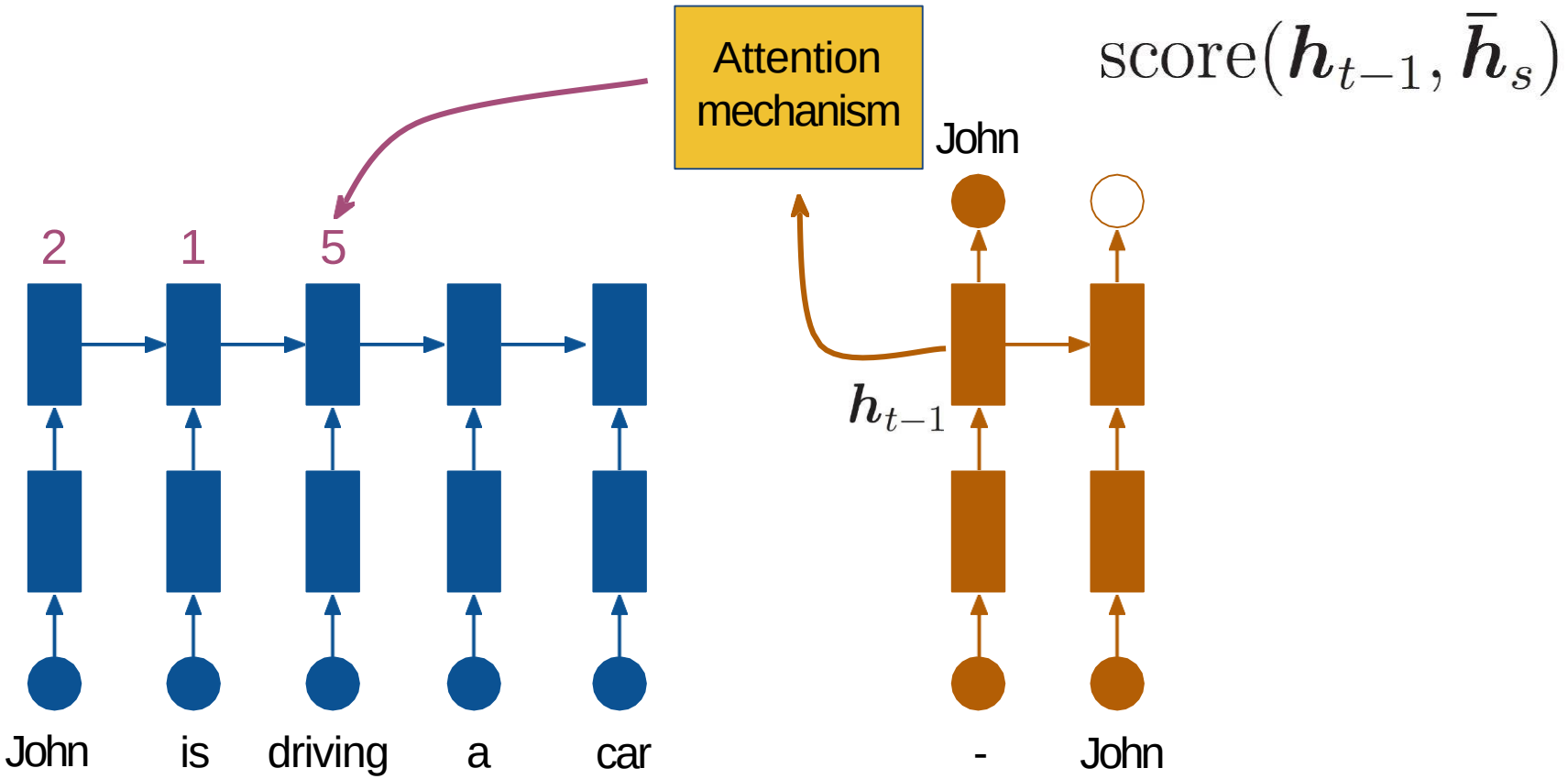
Neural Machine Translation by Luong, Cho and Manning

Attention Mechanism



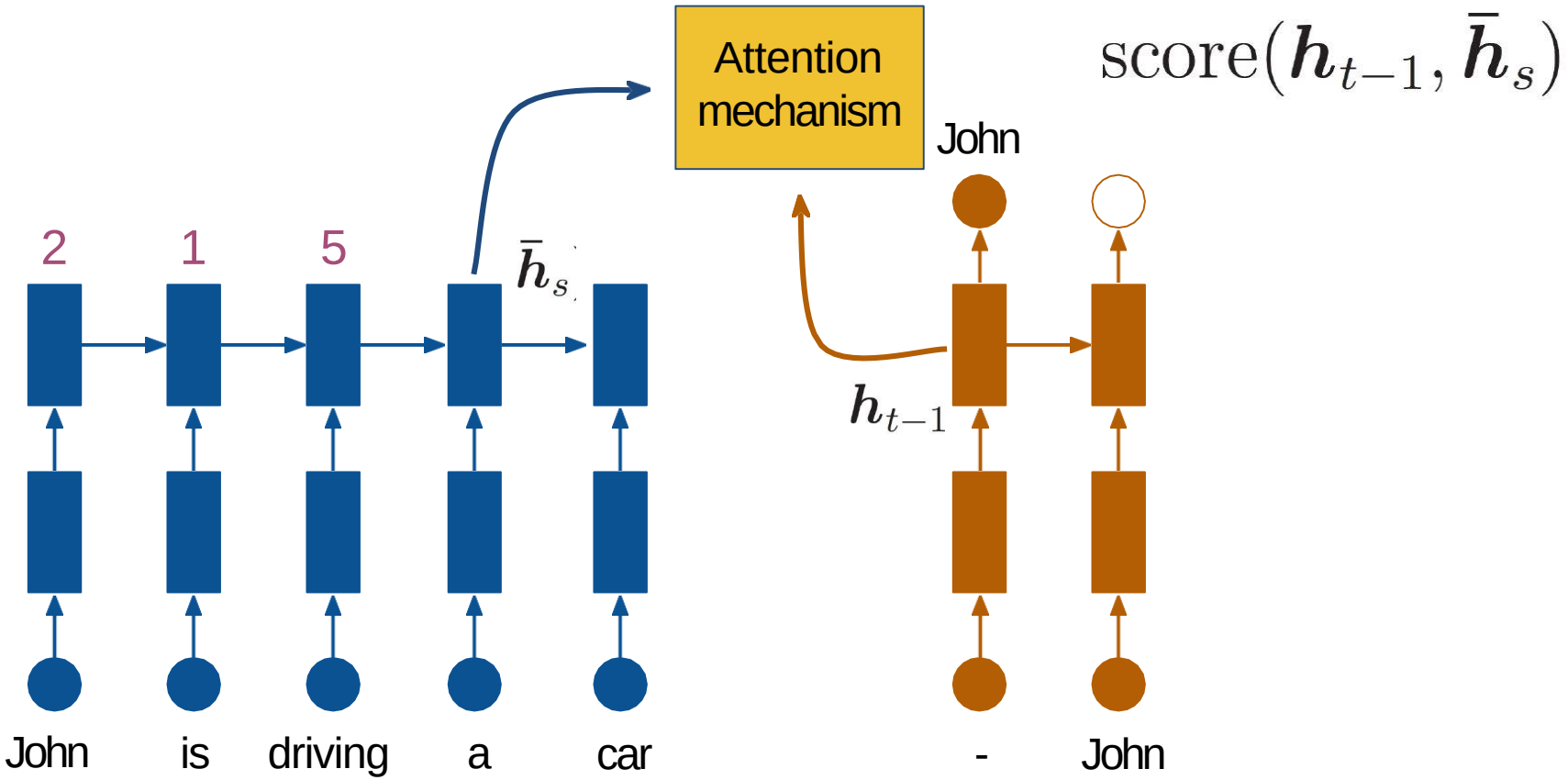
Neural Machine Translation by Luong, Cho and Manning

Attention Mechanism



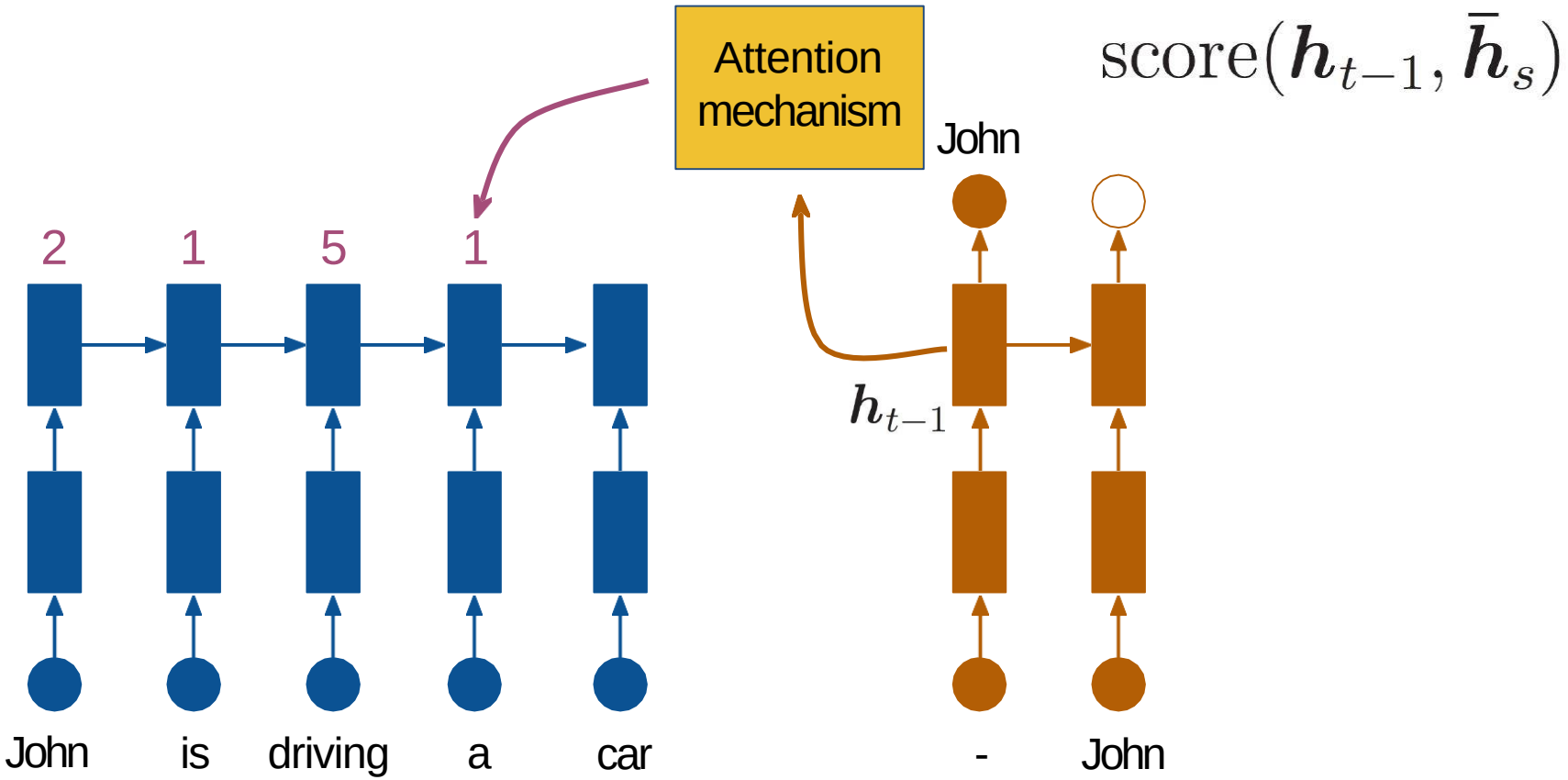
Neural Machine Translation by Luong, Cho and Manning

Attention Mechanism



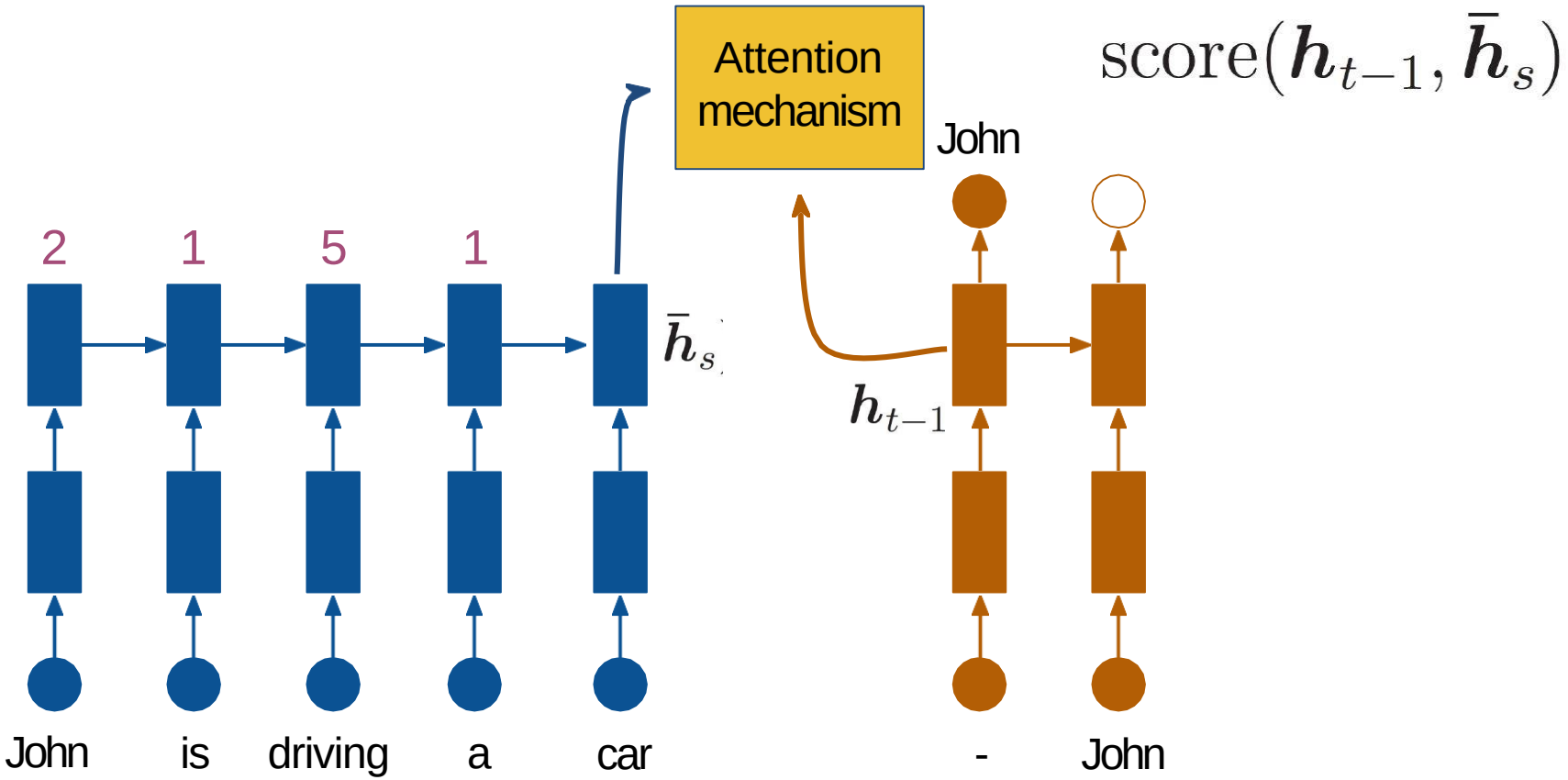
Neural Machine Translation by Luong, Cho and Manning

Attention Mechanism



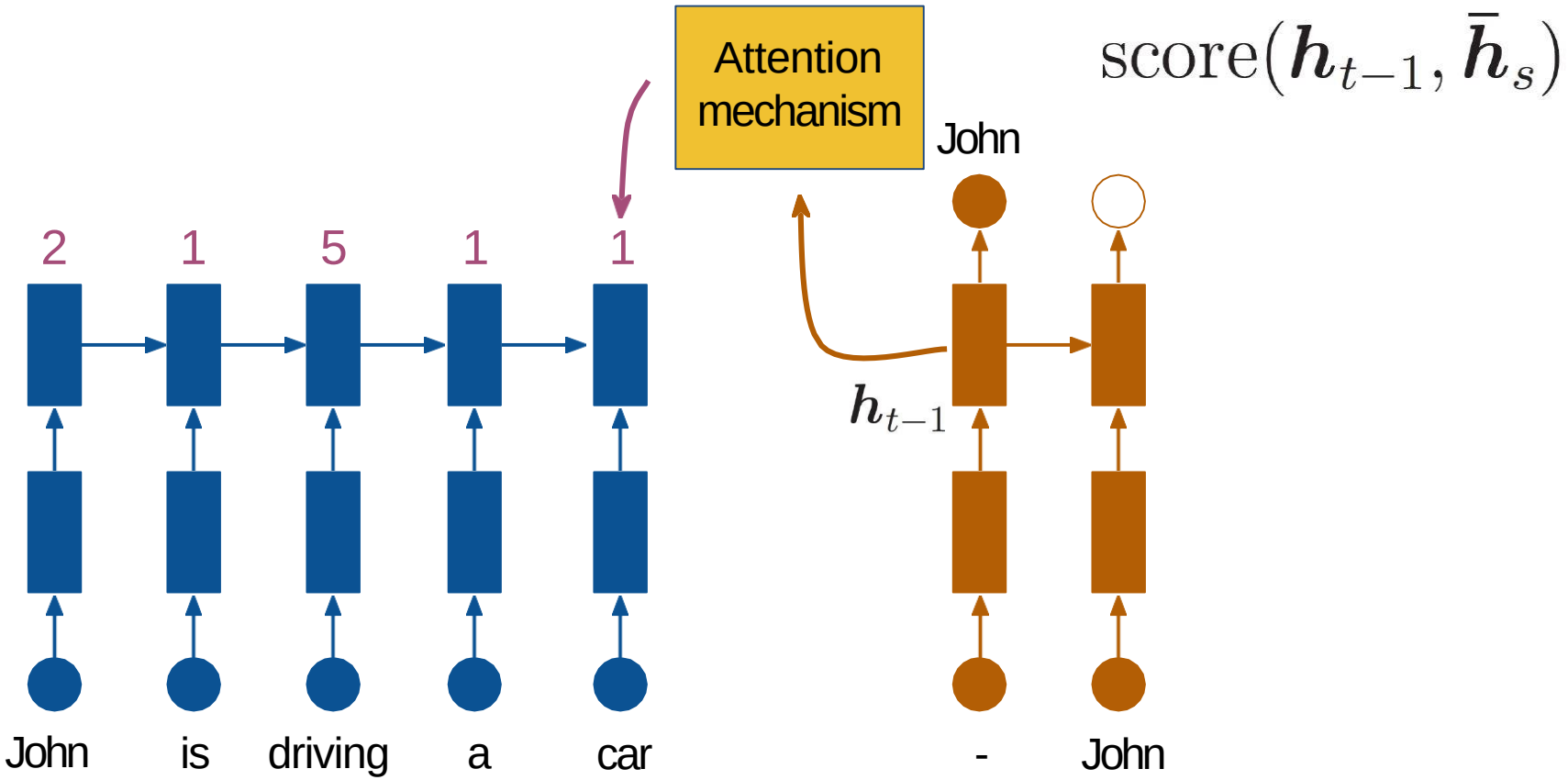
Neural Machine Translation by Luong, Cho and Manning

Attention Mechanism



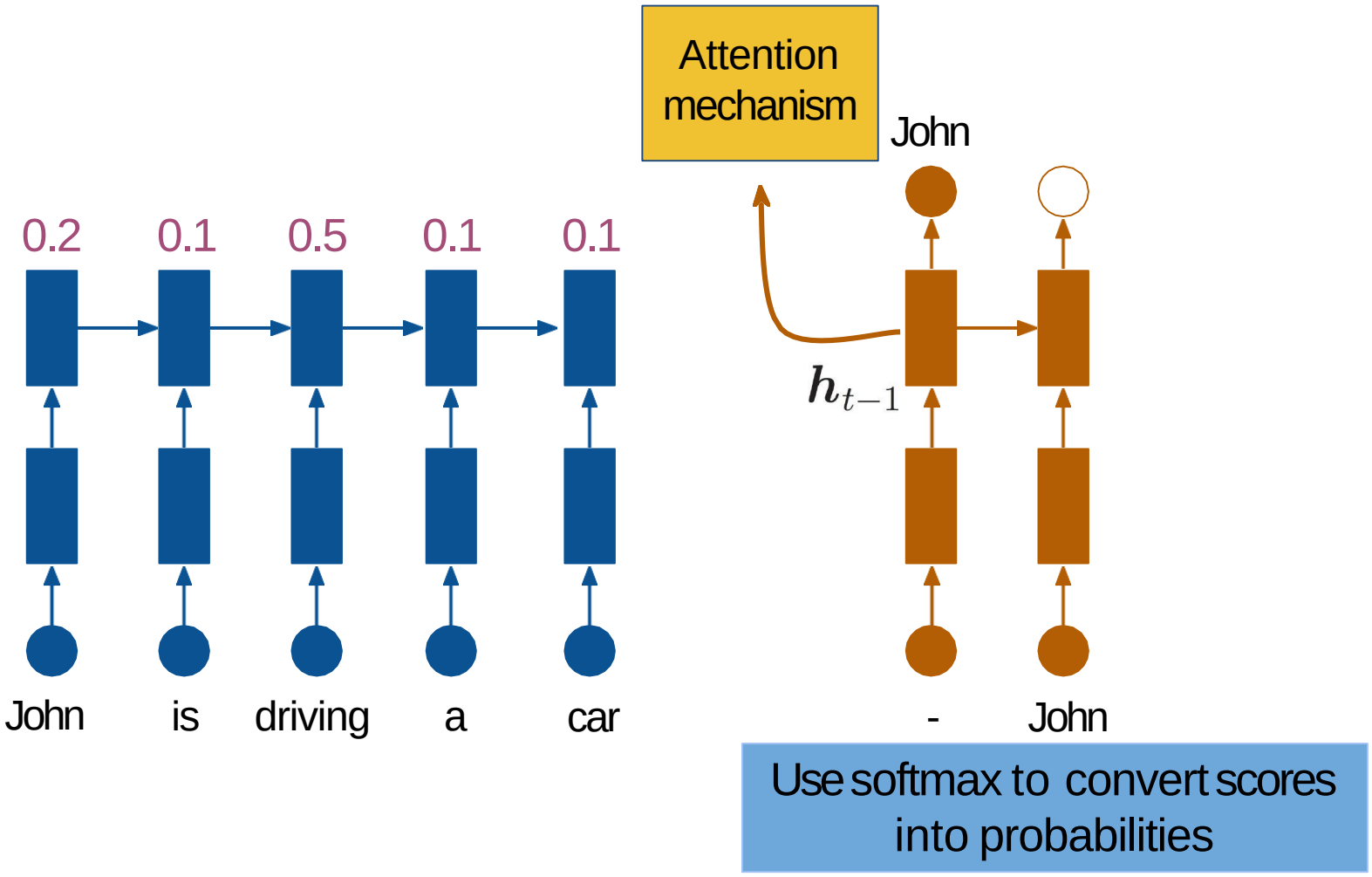
Neural Machine Translation by Luong, Cho and Manning

Attention Mechanism



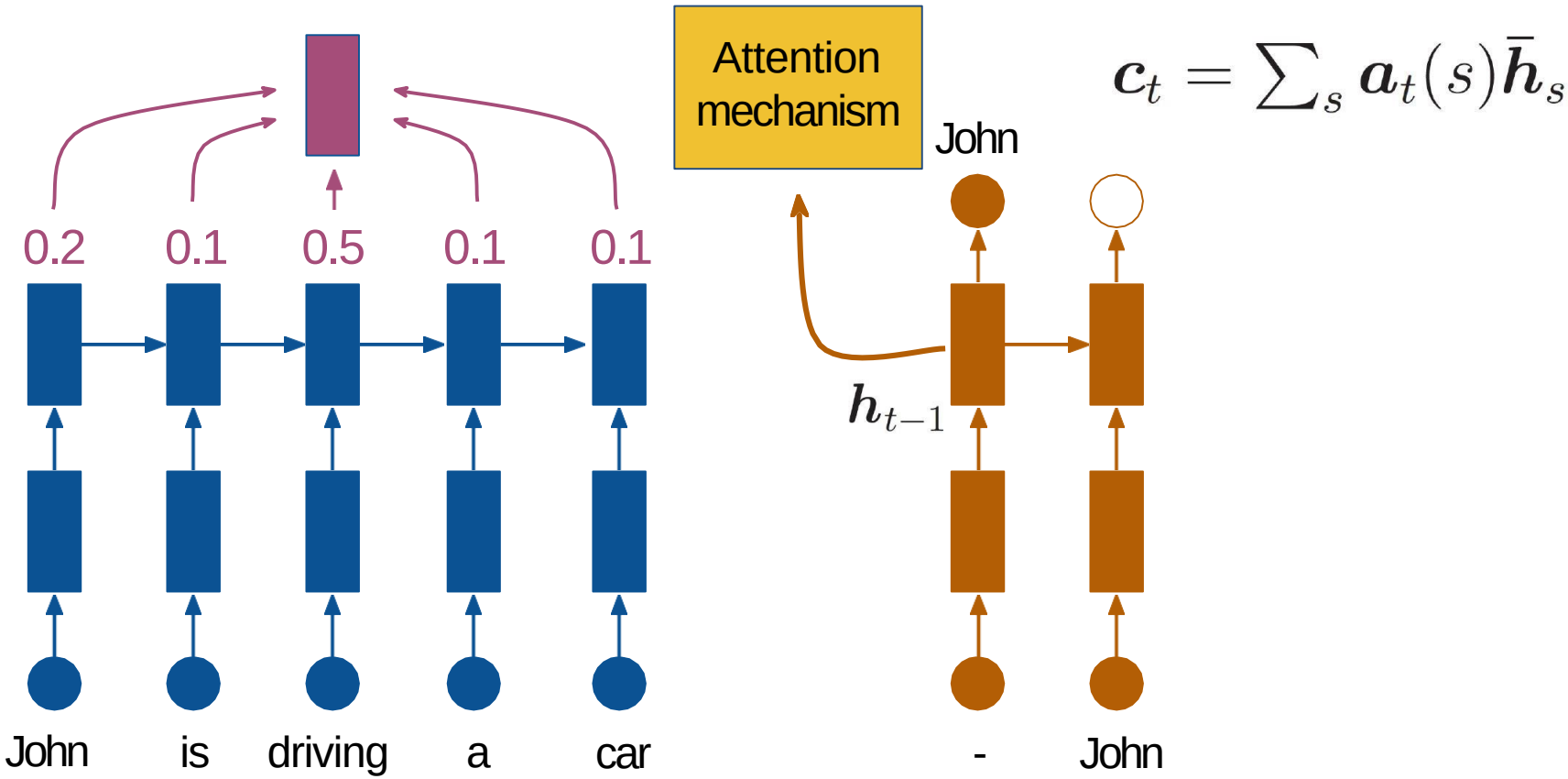
Neural Machine Translation by Luong, Cho and Manning

Attention Mechanism



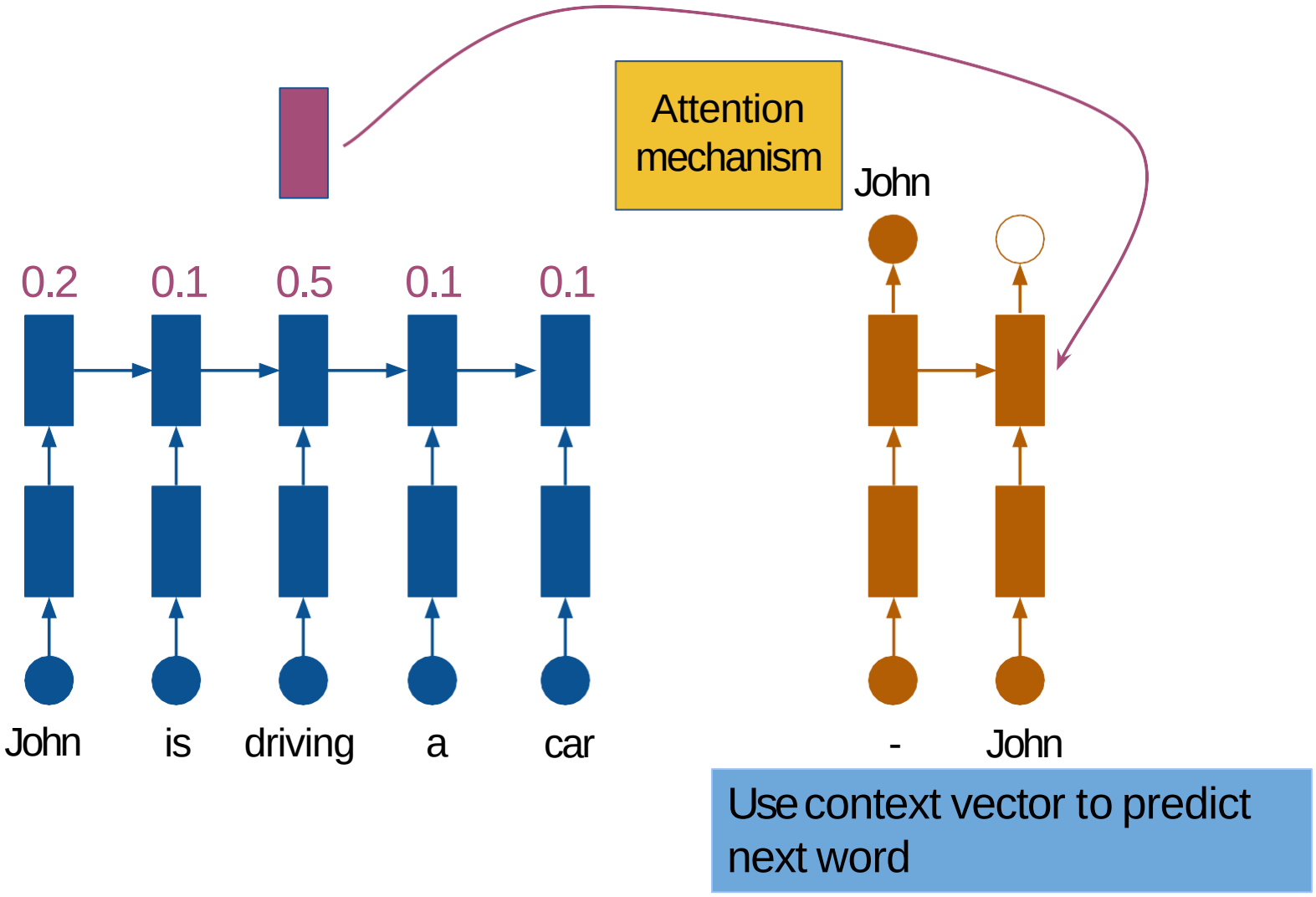
Neural Machine Translation by Luong, Cho and Manning

Attention Mechanism



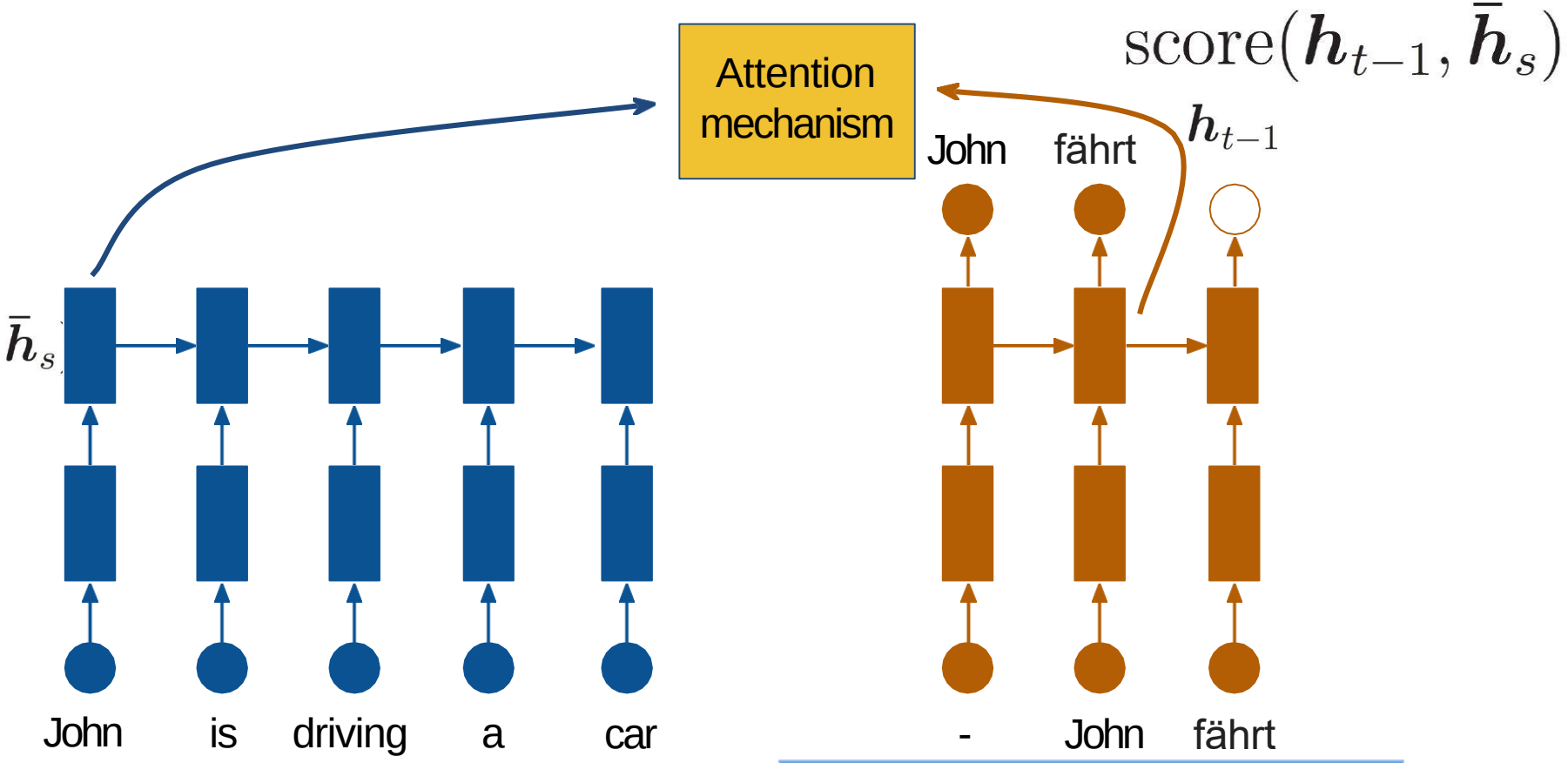
Neural Machine Translation by Luong, Cho and Manning

Attention Mechanism



Neural Machine Translation by Luong, Cho and Manning

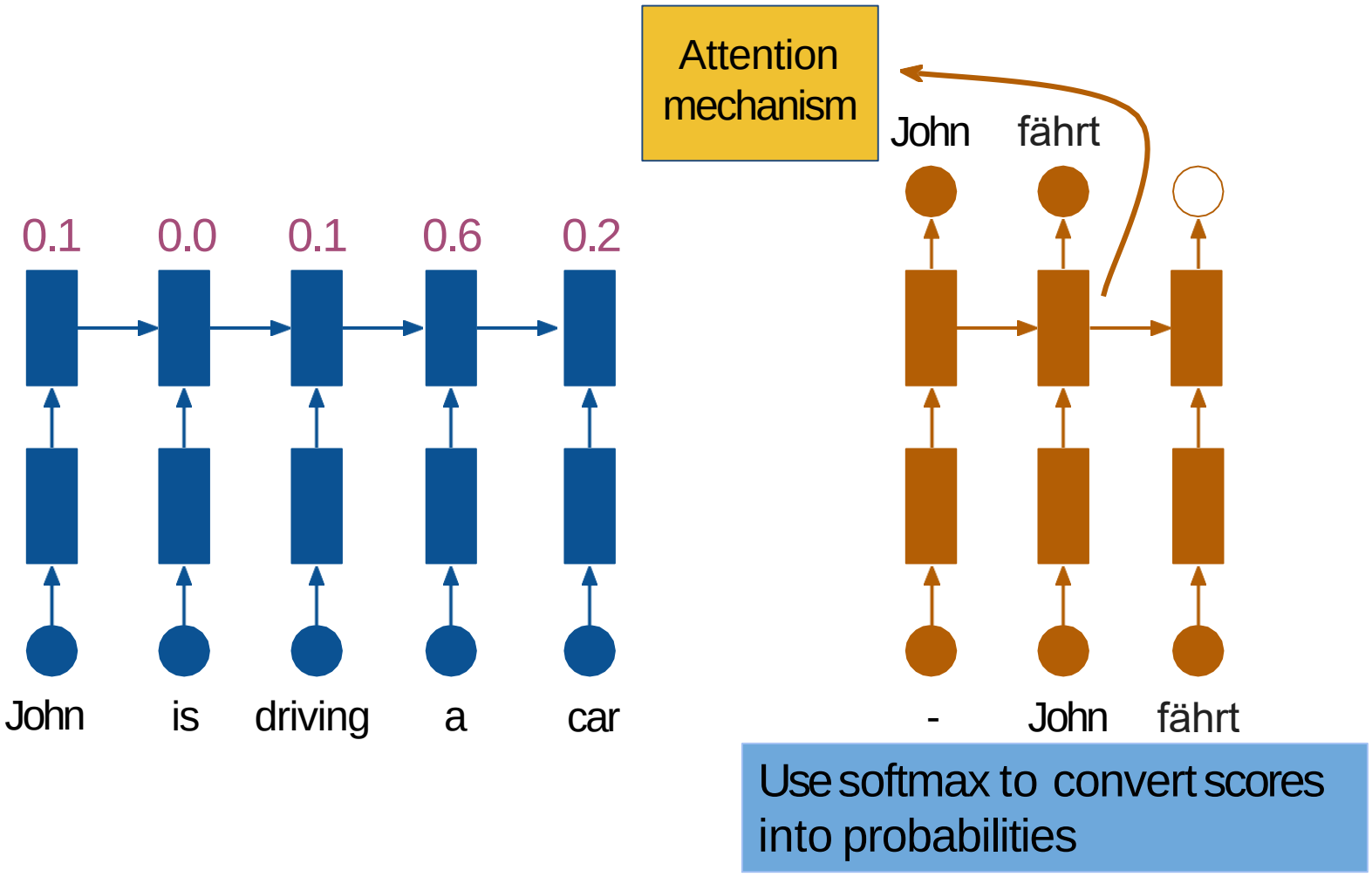
Attention Mechanism



Similarly, get *source scores* for the next prediction

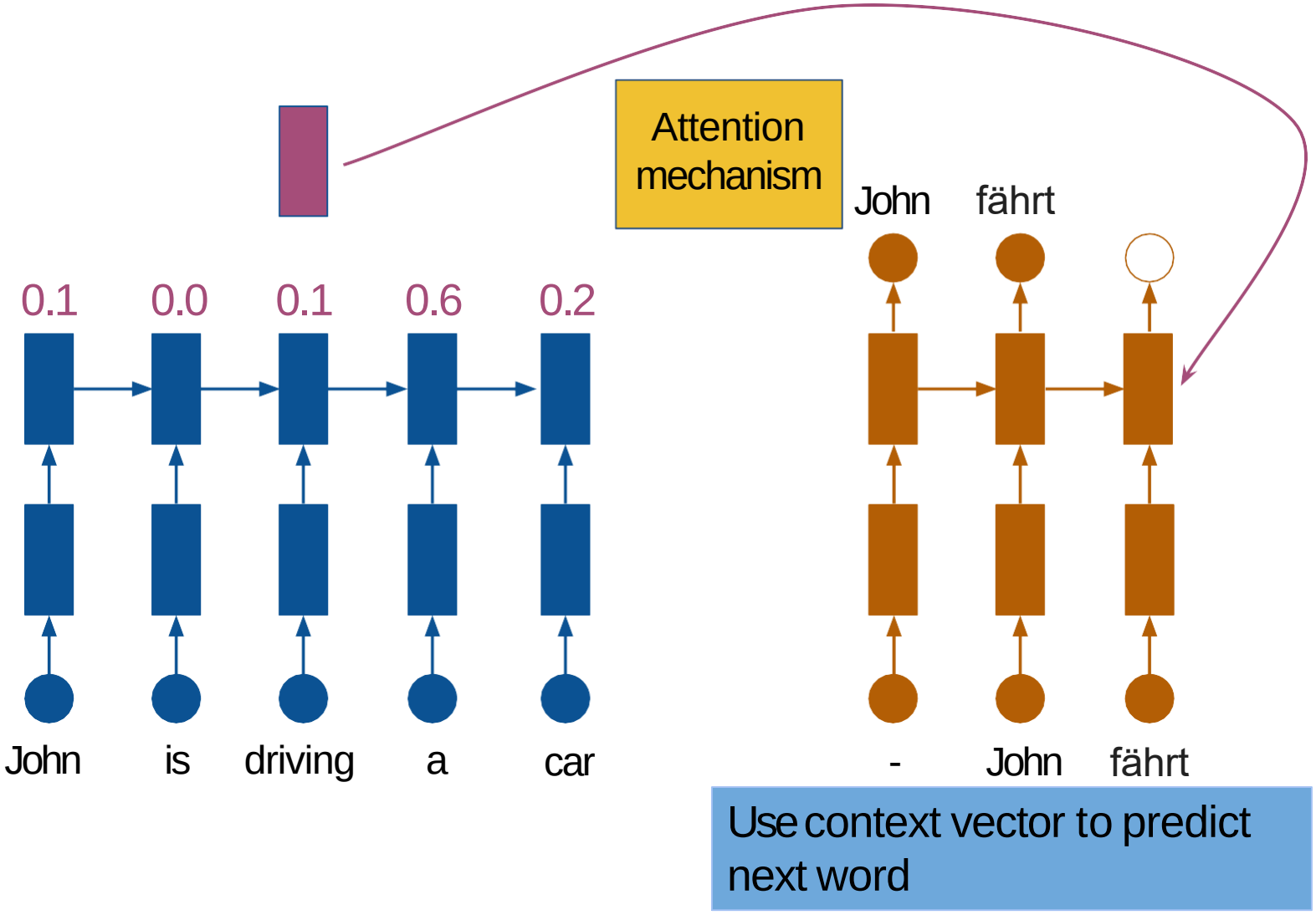
Neural Machine Translation by Luong, Cho and Manning

Attention Mechanism



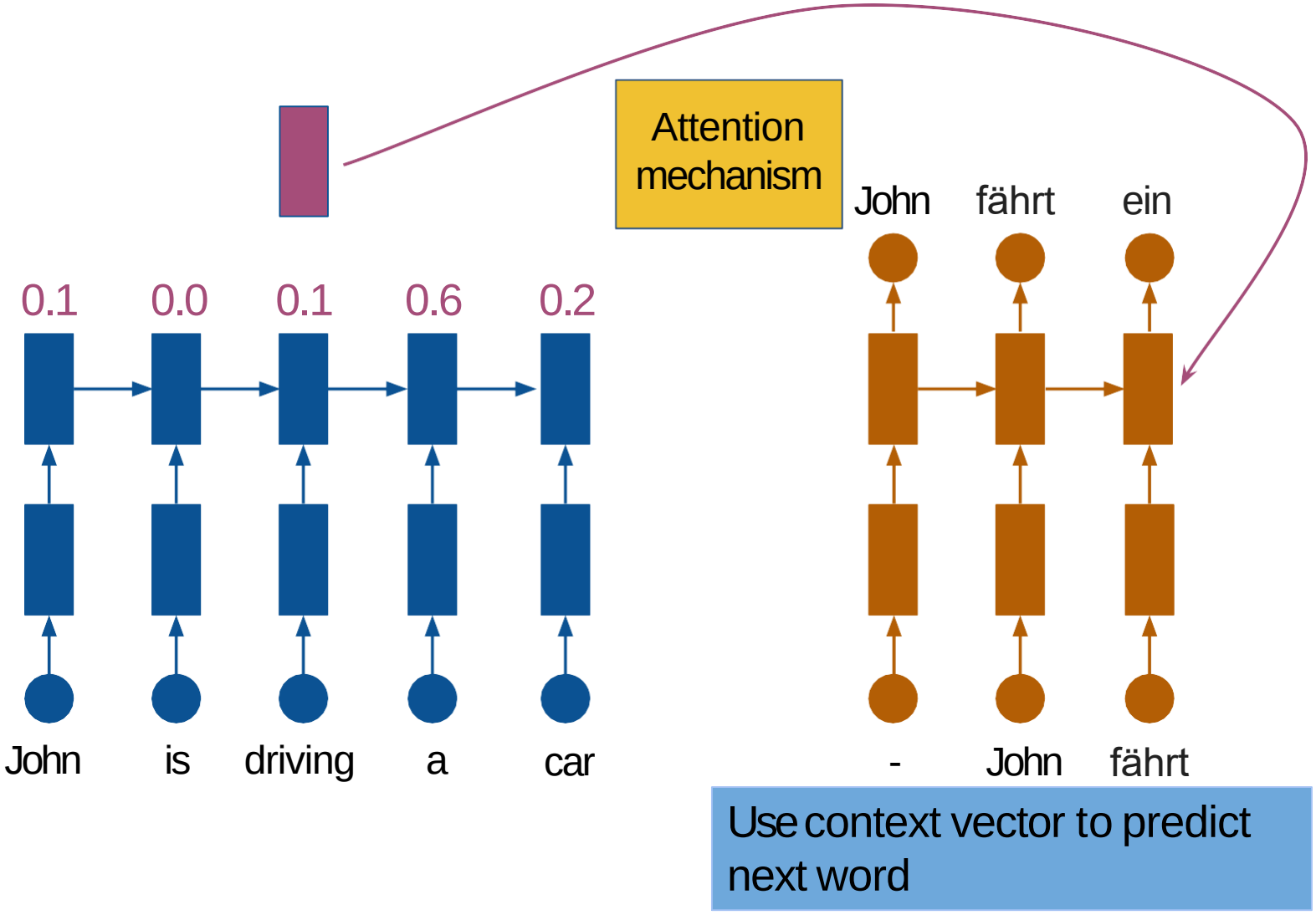
Neural Machine Translation by Luong, Cho and Manning

Attention Mechanism



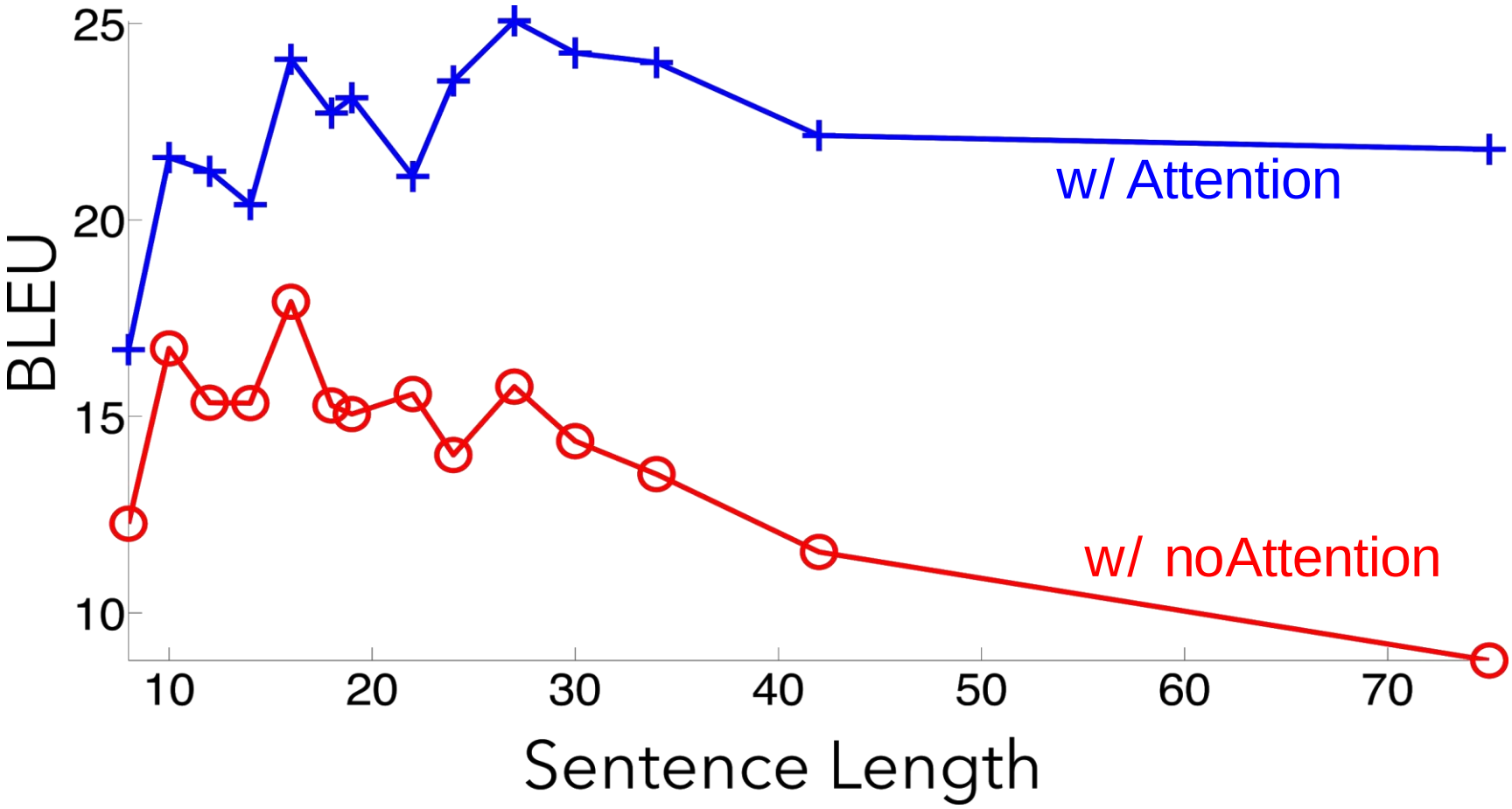
Neural Machine Translation by Luong, Cho and Manning

Attention Mechanism



Neural Machine Translation by Luong, Cho and Manning

Empirical Evaluation of Attention

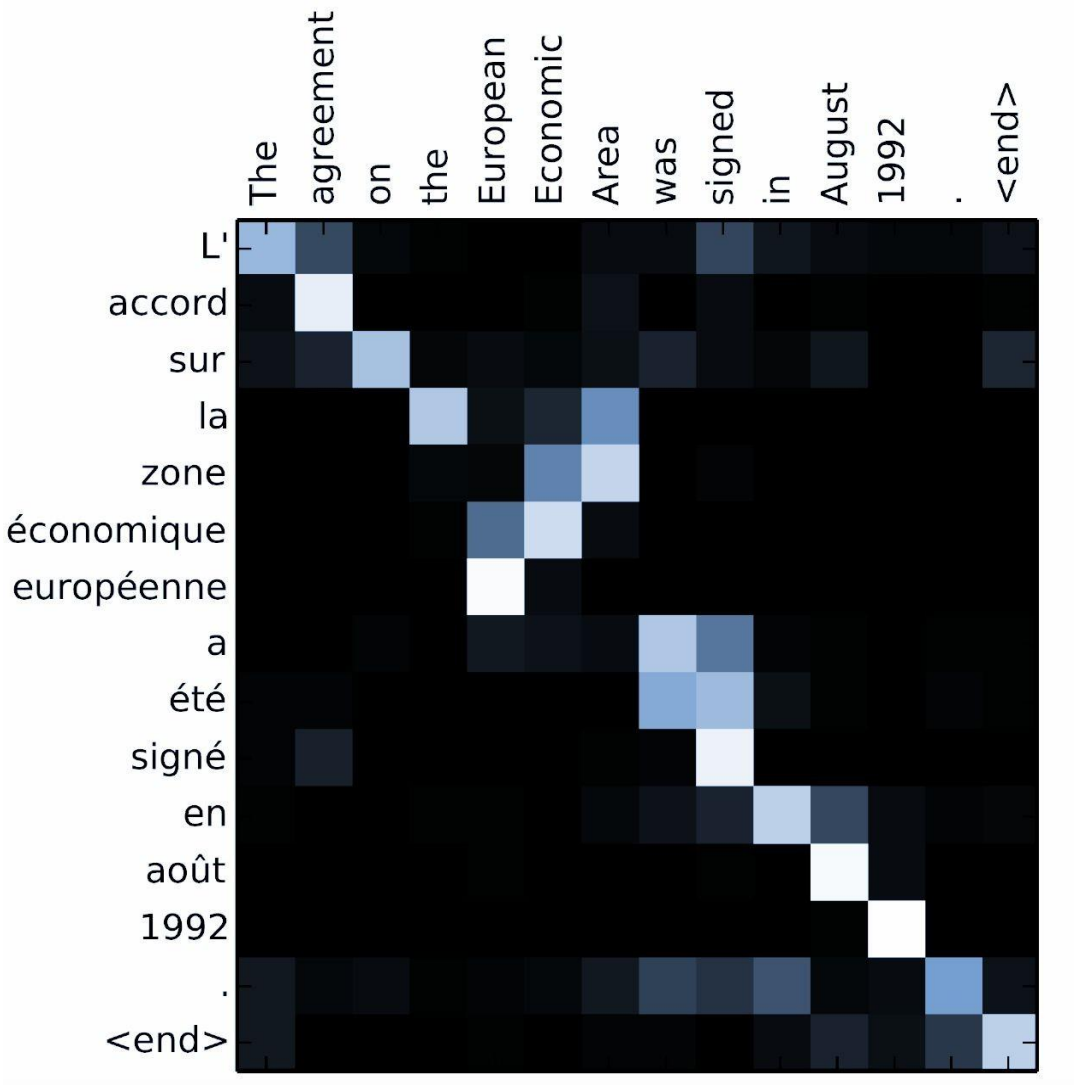


Neural Machine Translation by Luong, Cho and Manning

Attention Mechanism - Visual

Aligning words

If we plot the *source scores* for each target word, we can see what each target word is aligned to.



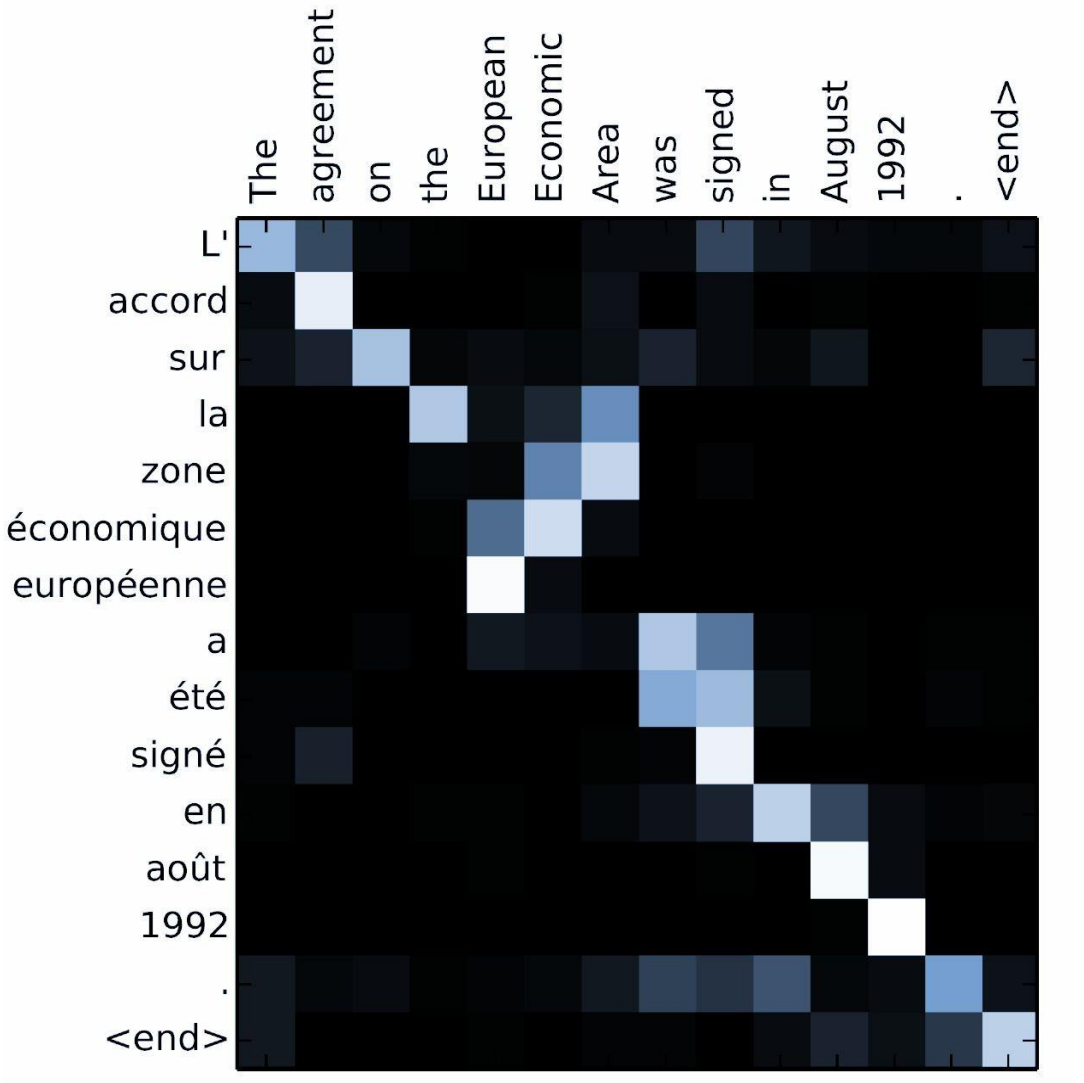
Dzmitry Bahdanau, KyungHuyn Cho, and Yoshua Bengio. Neural Machine Translation by Jointly Learning to Translate and Align. ICLR'15

Attention Mechanism - Visual

Aligning words

If we plot the *source scores* for each target word, we can see what each target word is aligned to.

Note the reordering in this particular example



Dzmitry Bahdanau, KyungHuyn Cho, and Yoshua Bengio. Neural Machine Translation by Jointly Learning to Translate and Align. ICLR'15

Applications of Sequence to Sequence Model

- Summarization
- Dialog-based systems (chatbots)
- Speech recognition
- Image captioning
- Machine translation
- ...

Image Captioning

"man in black shirt is playing guitar."

"construction worker in orange safety vest is working on road."

"two young girls are playing with lego toy."

https://towardsdatascience.com/image-captioning-in-deep-learning-9cd23fb4d8d2

Summary

- Bilingual LSTM translates from one language to another language
- Encoder-Decoder model helps us *encode* a sequence into a summary vector and use that to *decode* another sequence
- Attention mechanism learns soft alignment between source and target words
- OpenNMT toolkit http://opennmt.net/