RNNs

For different Tasys

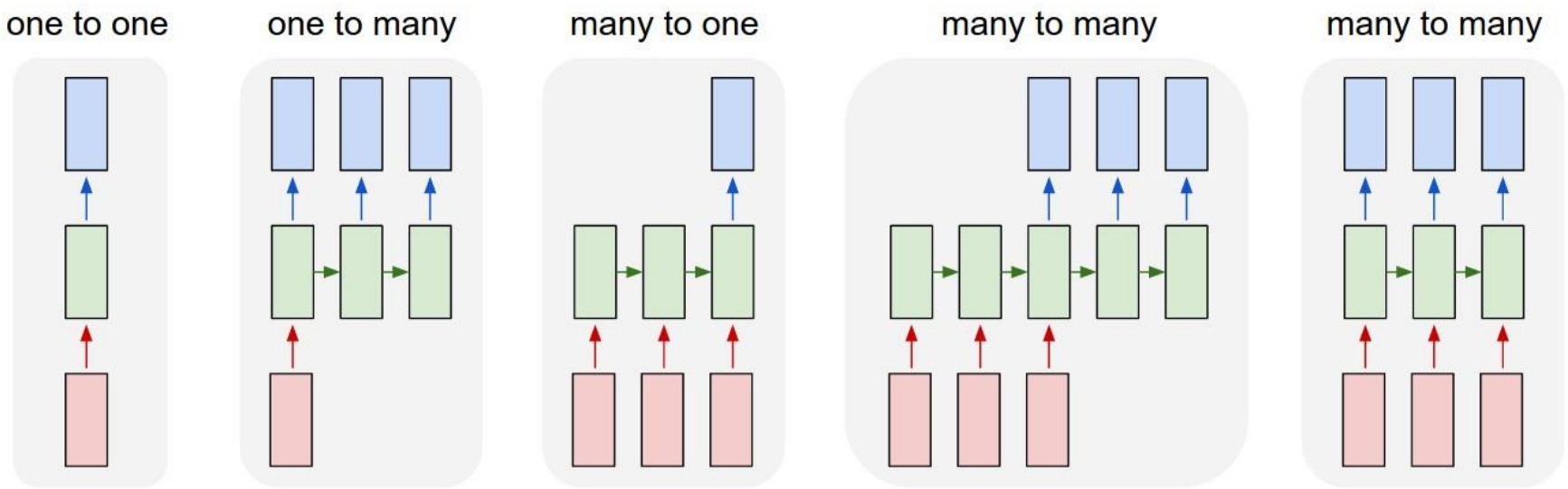
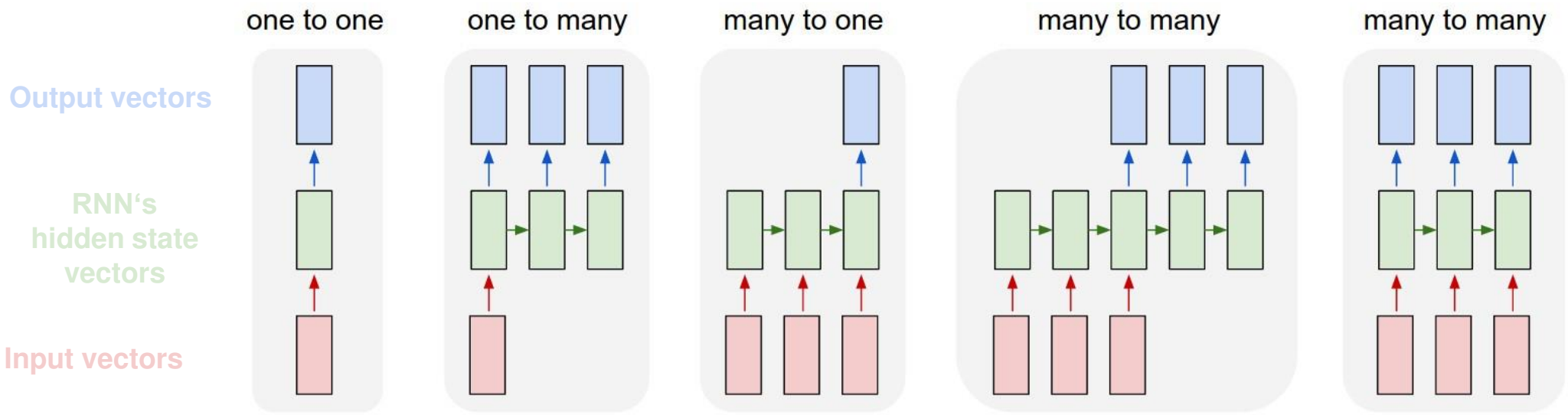
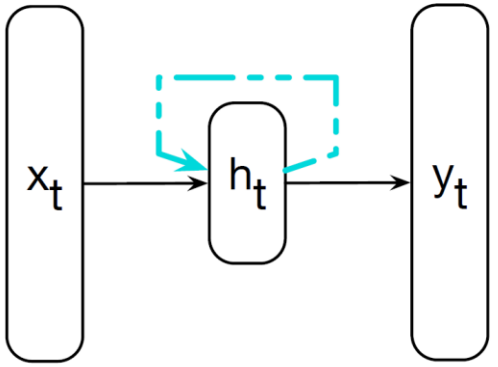


Image from: httx://karpathy.github.io/2015/05/21/rnn-effectiveness/

RNNs

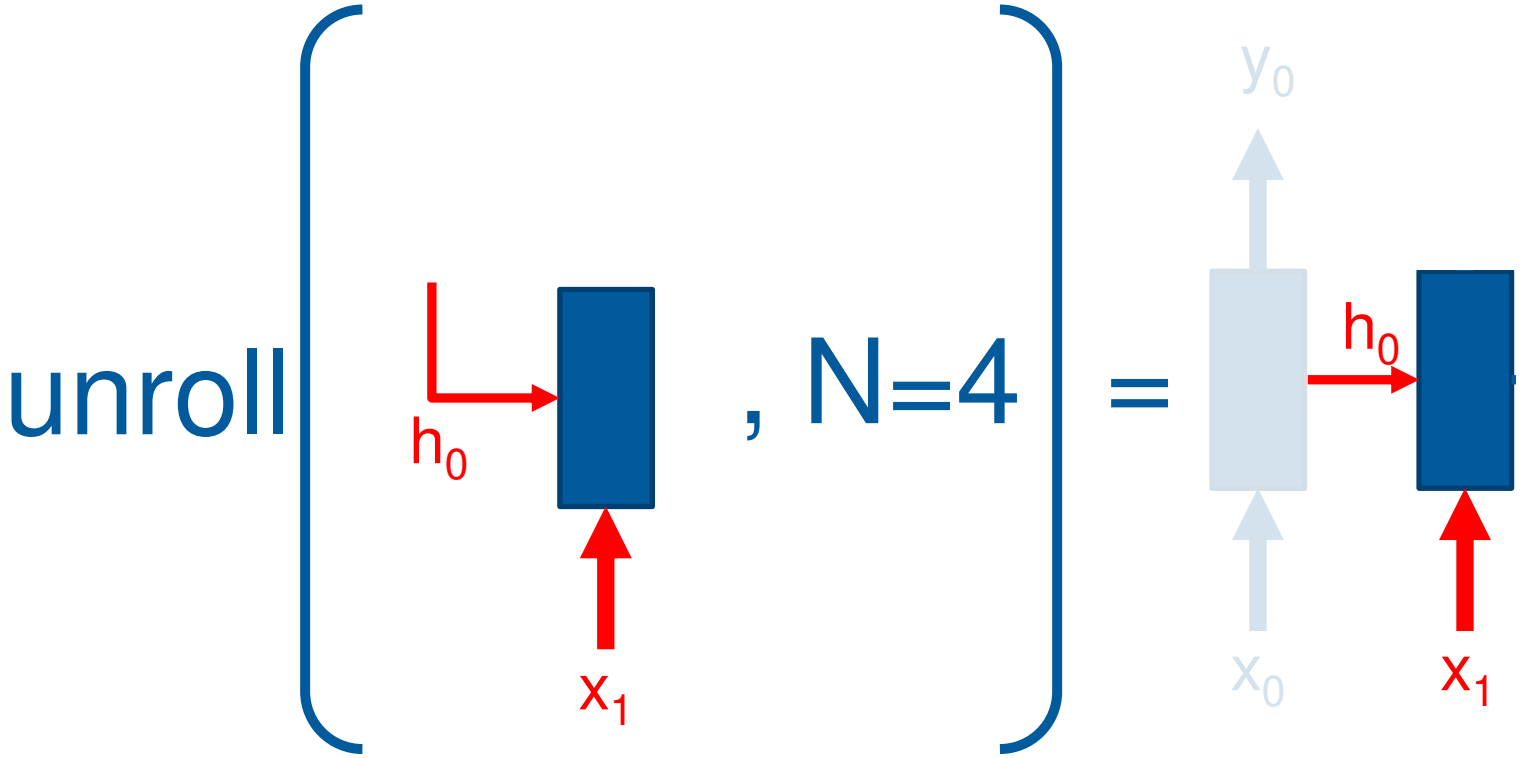
For different Tasys



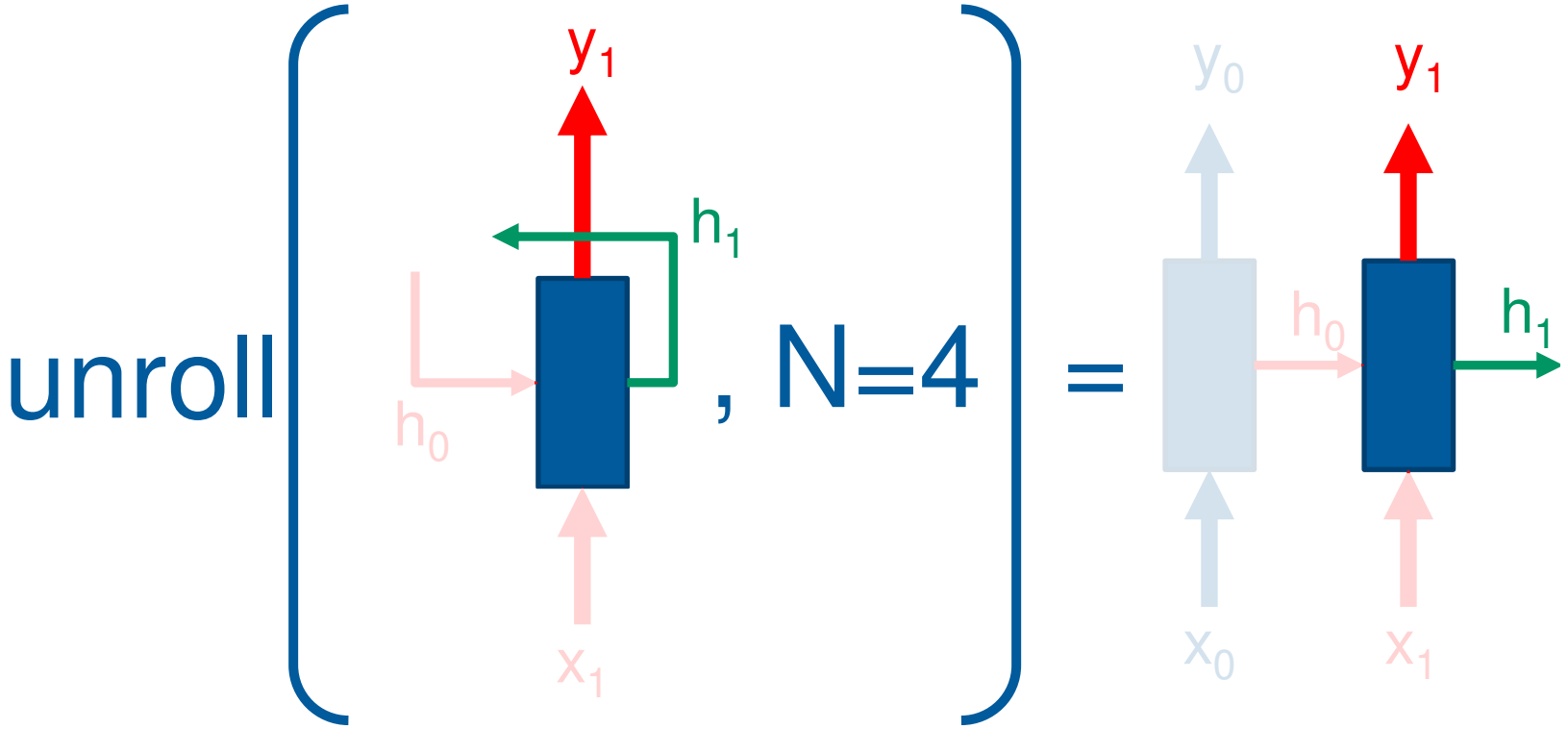


More about RNNs in general: Jurafsky, Section 9.1
RNNs and different applications: Jurafsky, 9.3
Image from: httx://karpathy.github.io/2015/05/21/rnn-effectiveness/

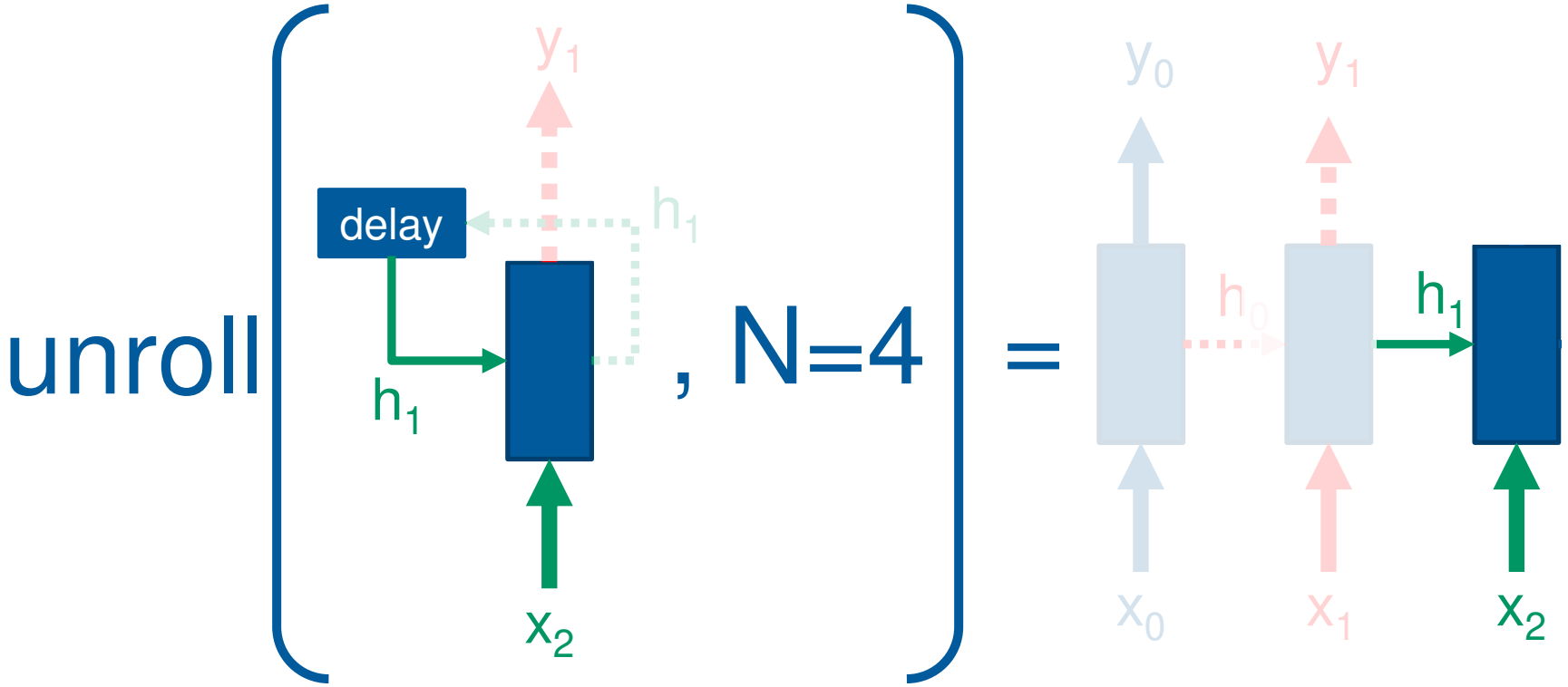
RNN and unrolling of RNN



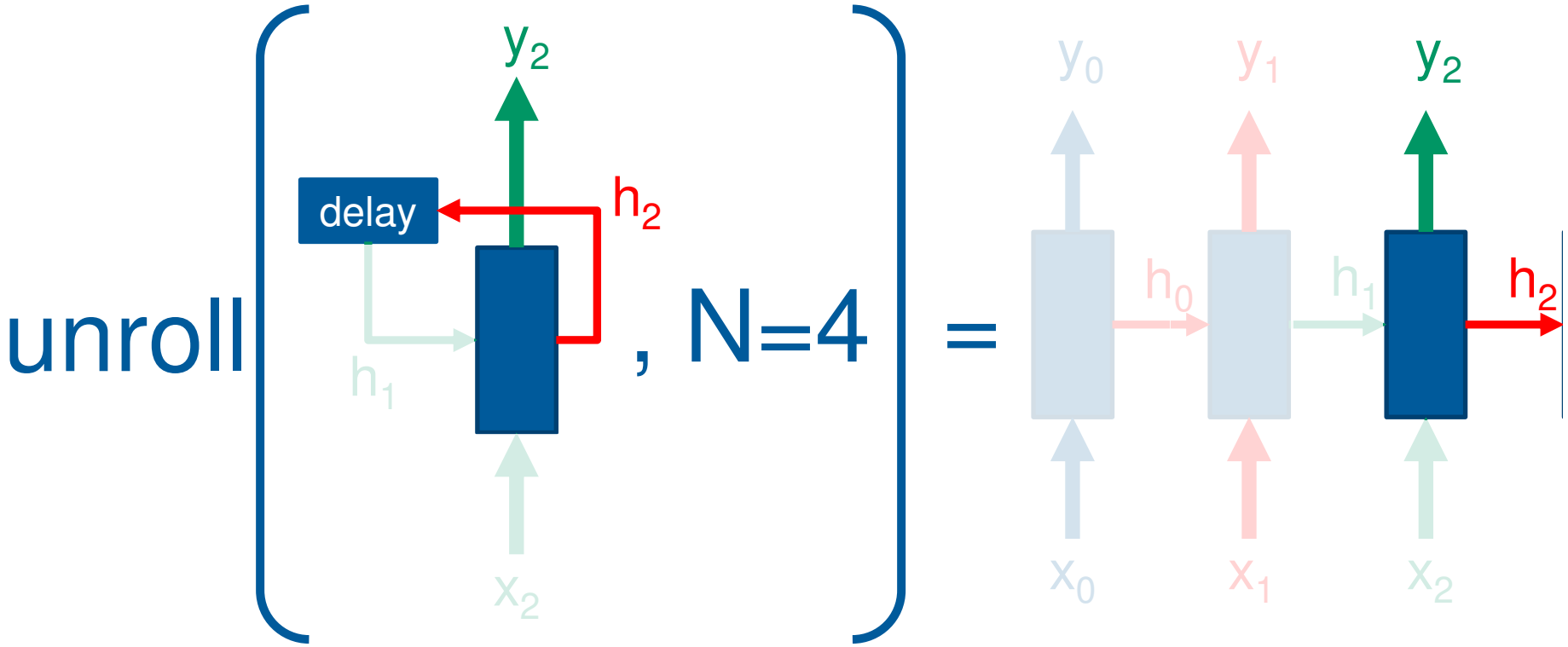
RNN and unrolling of RNN



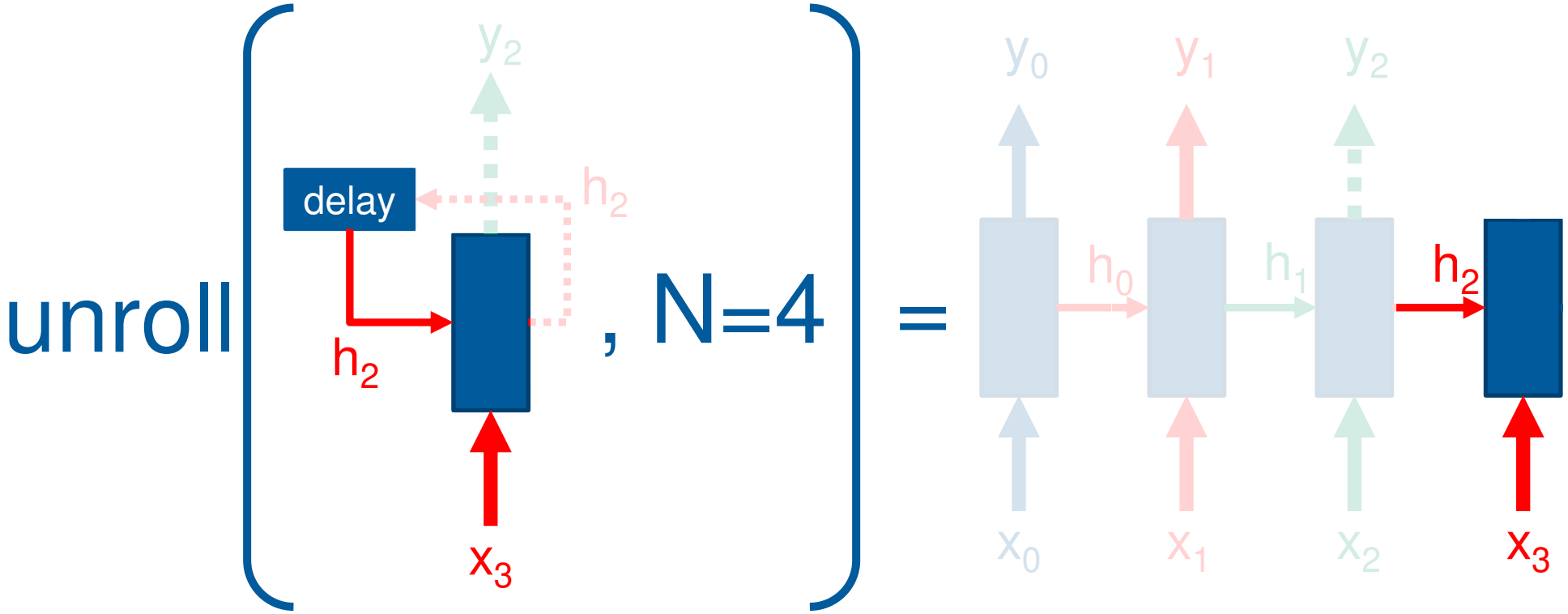
RNN and unrolling of RNN



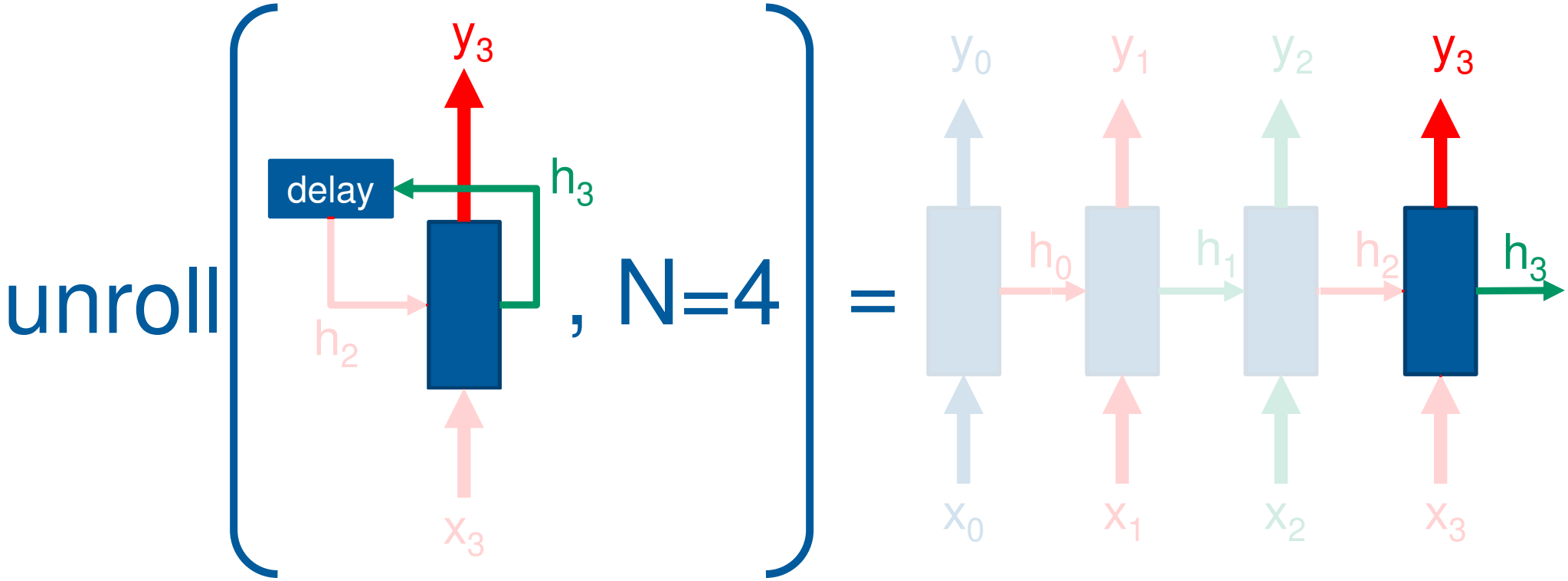
RNN and unrolling of RNN



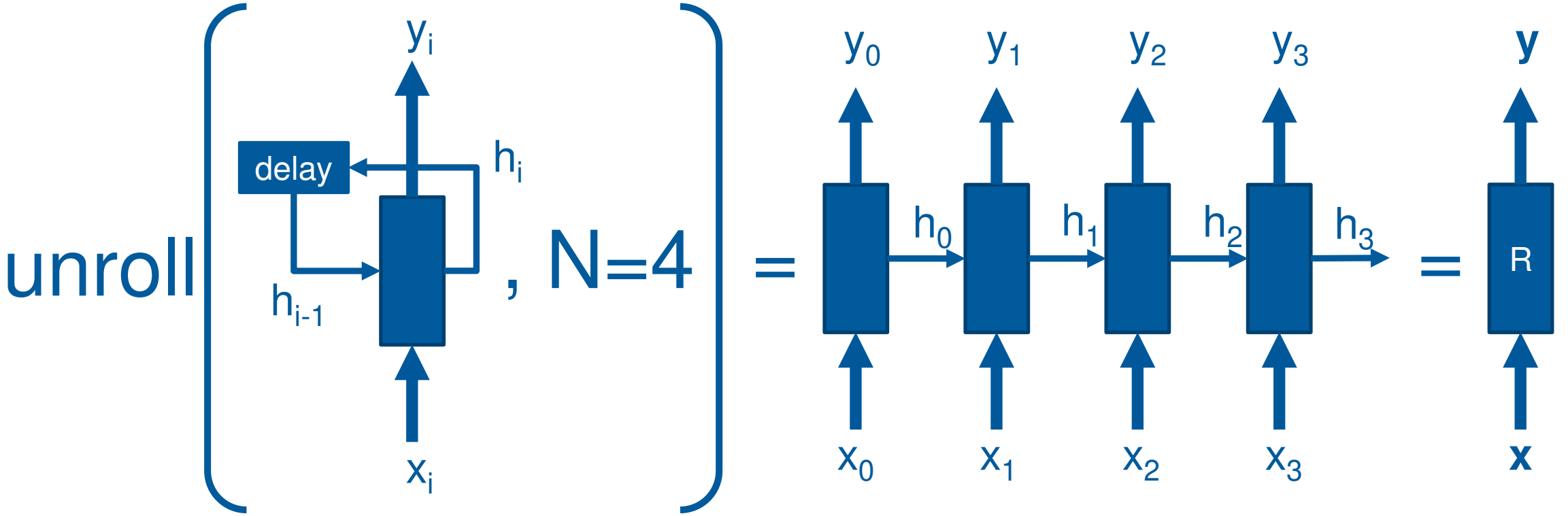
RNN and unrolling of RNN



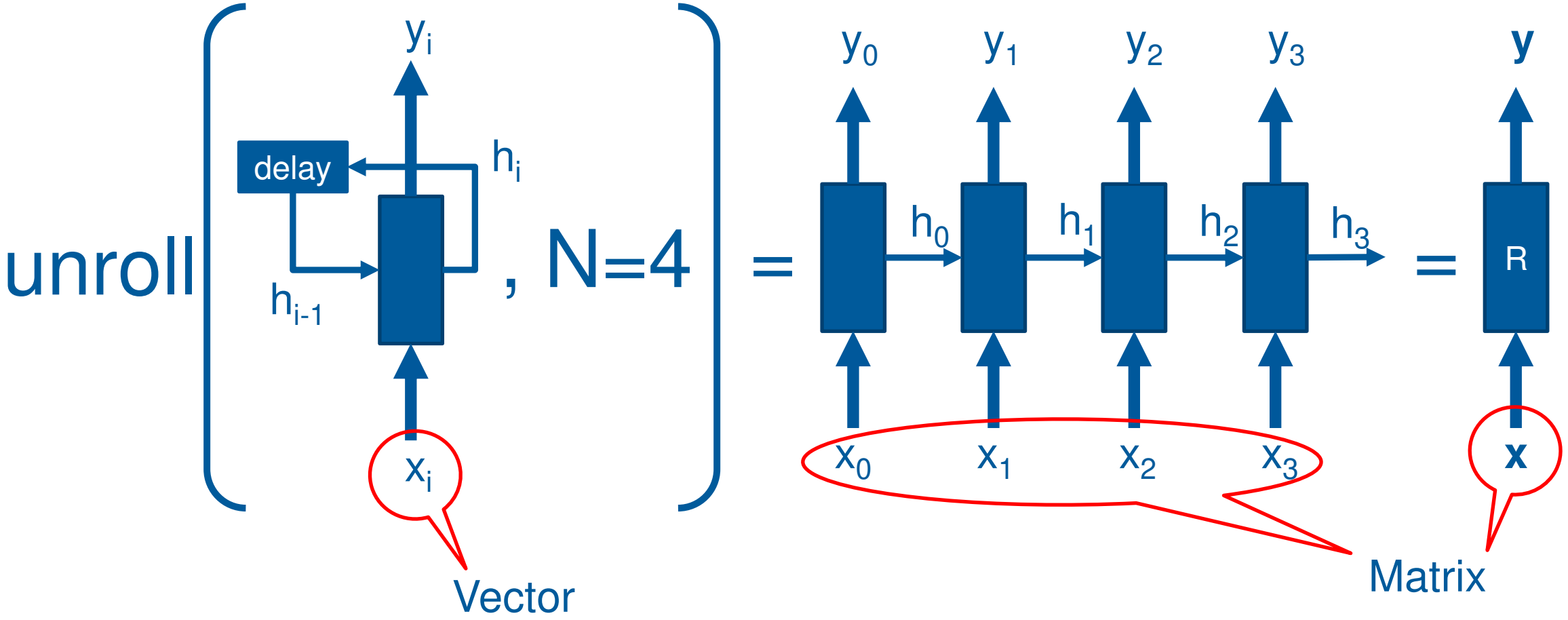
RNN and unrolling of RNN



RNN and unrolling of RNN



RNN and unrolling of RNN



https://machinelearningmastery.com/rnn-unrolling/

Named Entity Recognition or more general “Slot Extraction”

Used in Personal Assistants but also in information mining

Seeks to locate and classify named entities mentioned in unstructured text into pre-defined categories such as person names, organizations, locations, medical codes, time expressions, quantities, monetary values, percentages, etc.

User: PLEASE PLAY WE WILL ROCK YOU BY QUEEN

System:

Title: <Play we will Rock You>

Artist: <Queen>

User: REDUCE BRIGHTESS TO LEVEL FIVE

System:

Level: <5>

https://arxiv.org/pdf/2003.04807.pdf

Named Entity Recognition or more general “Slot Extraction”

Formalization

Seeks to locate and classify named entities mentioned in unstructured text into pre-defined categories such as person names, organizations, locations, medical codes, time expressions, quantities, monetary values, percentages, etc.

User: PLEASE PLAY WE WILL ROCK YOU BY QUEEN

System:

Title: <Play we will Rock You>

Artist: <Queen>

$$\hat{s}_t = \underset{s \in \text{Tags}}{\text{argmax}}\, P(s|w_t)$$

User: REDUCE BRIGHTESS TO LEVEL FIVE

System:

Level: <5>

https://arxiv.org/pdf/2003.04807.pdf

Named Entity Recognition or more general “Slot Extraction”

Recurrent Neural Networks

Seeks to locate and classify named entities mentioned in unstructured text into pre-defined categories such as person names, organizations, locations, medical codes, time expressions, quantities, monetary values, percentages, etc.

User: PLEASE PLAY WE WILL ROCK YOU BY QUEEN

System:

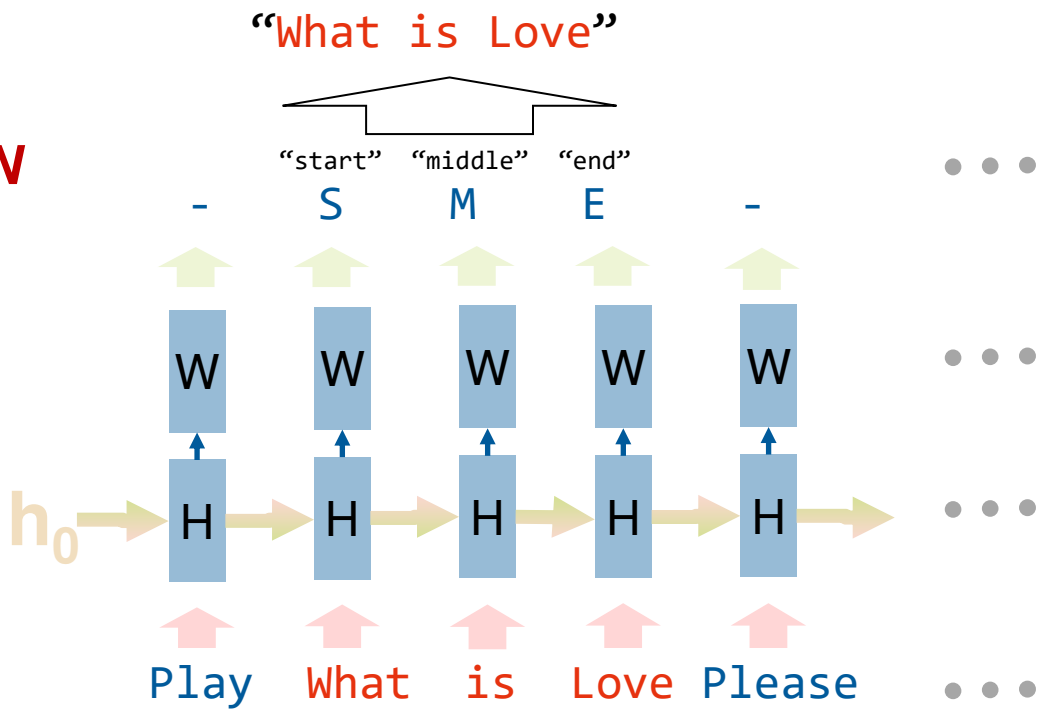
Title: <Play we will Rock You>

Artist: <Queen>

User: REDUCE BRIGHTESS TO LEVEL FIVE

System:

Level: <5>



https://arxiv.org/pdf/2003.04807.pdf

Named Entity Recognition or more general “Slot Extraction”

Recurrent Neural Networks

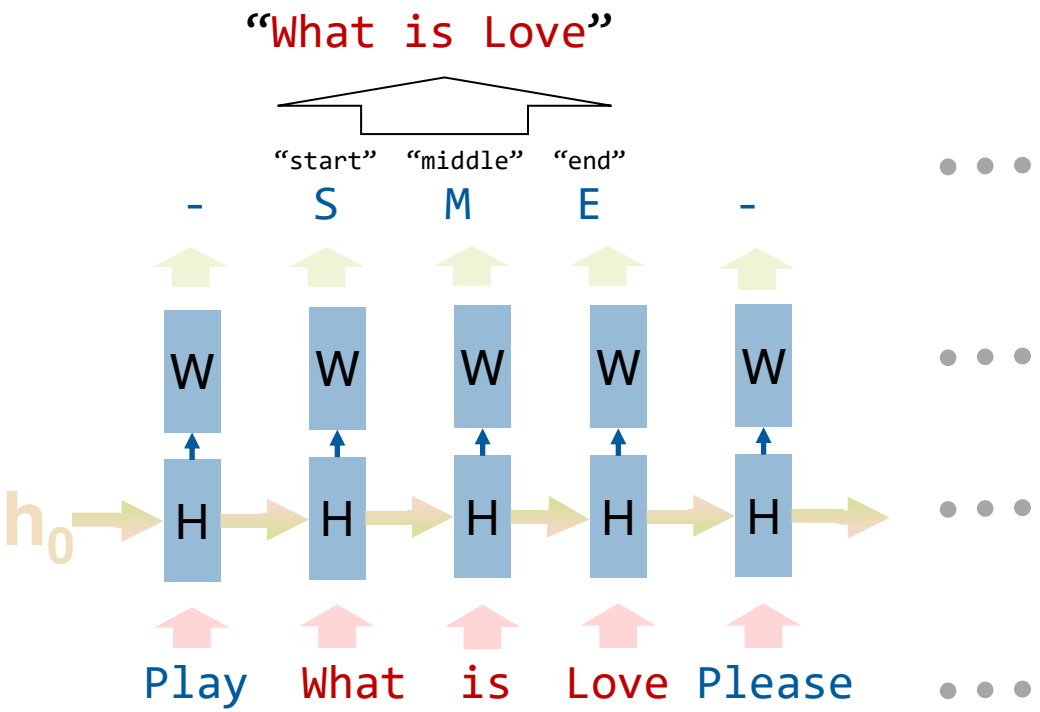
Seeks to locate and classify named entities mentioned in unstructured text into pre-defined categories such as person names, organizations, locations, medical codes, time expressions, quantities, monetary values, percentages, etc.

Word Embedding

$$x = [Ew_{i-n}, \ldots, Ew_i, \ldots, Ew_{i+n}]$$

$$h_t = \text{sigmoid}(W_x x + W_h h_{t-1} + b_h)$$

$$\hat{s}_t = \underset{s \in \text{Tags}}{\text{argmax}}(\text{softmax}(W h_t + b)_s)$$



https://arxiv.org/pdf/2003.04807.pdf

Sentimental Analysis (or opinion mining)

Recurrent Neuronal Networks

Sentiment analysis uses natural language processing to interpret and classify emotions in subjective data. Sentiment analysis is often used in business to detect sentiment in social data, gauge brand reputation, and understand customers.

User: It‘s a wonderful day.

- ***Sentiment: Positive***

User: It‘s a boring day.

- ***Sentiment: Negative***

Sentimental Analysis (or opinion mining)

Recurrent Neuronal Networks

Sentiment analysis uses natural language processing to interpret and classify emotions in subjective data. Sentiment analysis is often used in business to detect sentiment in social data, gauge brand reputation, and understand customers.

User: It‘s a wonderful day.

- ***Sentiment: Positive***

User: It‘s a boring day.

- ***Sentiment: Negative***

Sentimental Analysis (or opinion mining)

Recurrent Neuronal Networks

Sentiment analysis uses natural language processing to interpret and classify emotions in subjective data. Sentiment analysis is often used in business to detect sentiment in social data, gauge brand reputation, and understand customers.

User: It‘s a wonderful day.

- ***Sentiment: Positive***

$$\hat{s}_t = \underset{s \in \text{Tags}}{\text{argmax}}\, P(s|w_t)$$

User: It‘s a boring day.

- ***Sentiment: Negative***

Sentimental Analysis (or opinion mining)

Recurrent Neuronal Networks

Sentiment analysis uses natural language processing to interpret and classify emotions in subjective data. Sentiment analysis is often used in business to detect sentiment in social data, gauge brand reputation, and understand customers.

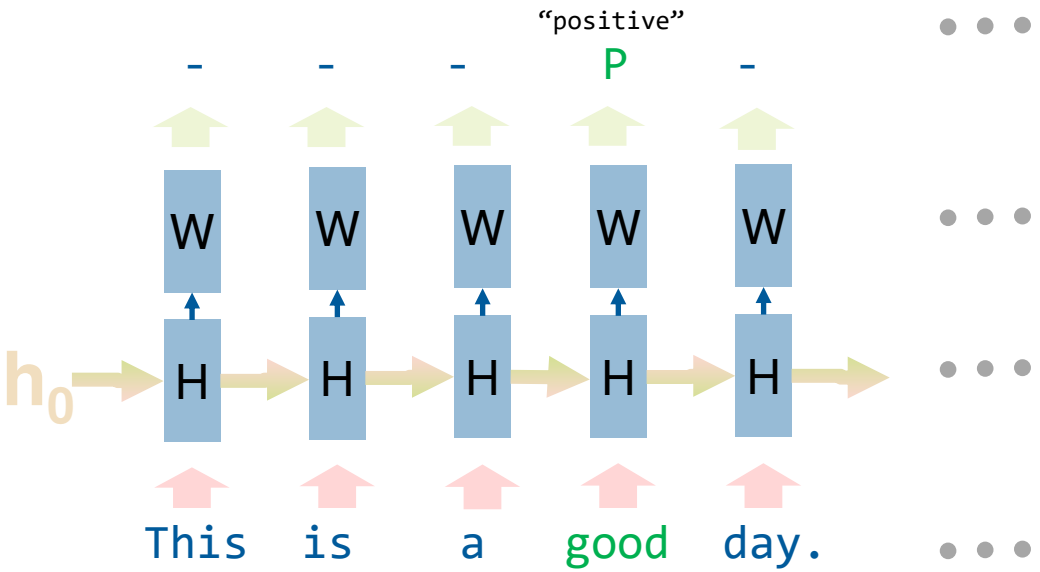
User: It's a wonderful day.

- ***Sentiment: Positive***

User: It's a boring day.

- ***Sentiment: Negative***

$$\hat{s}_t = \underset{s \in \text{Tags}}{\text{argmax}}\ P(s|w_t)$$



Sentimental Analysis (or opinion mining)

Recurrent Neuronal Networks

Sentiment analysis uses natural language processing to interpret and classify emotions in subjective data. Sentiment analysis is often used in business to detect sentiment in social data, gauge brand reputation, and understand customers.

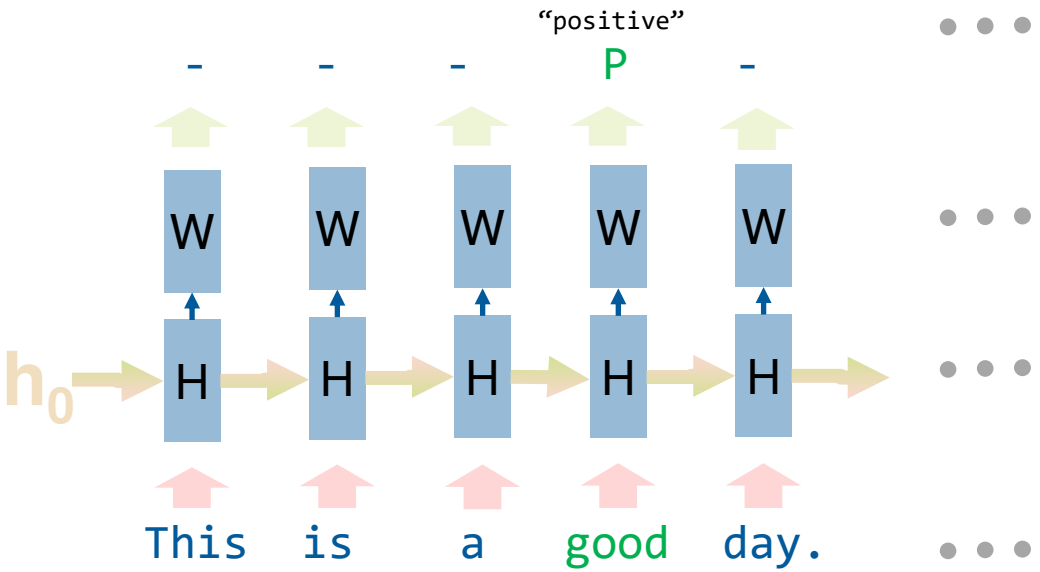
$$\hat{s}_t = \underset{s \in \text{Tags}}{\text{argmax}}\, P(s|w_t)$$

Word Embedding

$$x = [Ew_{i-n}, \ldots, Ew_i, \ldots, Ew_{i+n}]$$

$$h_t = \text{sigmoid}(W_x x + W_h h_{t-1} + b_h)$$

$$\hat{s}_t = \underset{s \in \text{Tags}}{\text{argmax}}(\text{softmax}(W h_t + b)_s)$$



Sentimental Analysis (or opinion mining)

Recurrent Neuronal Networks

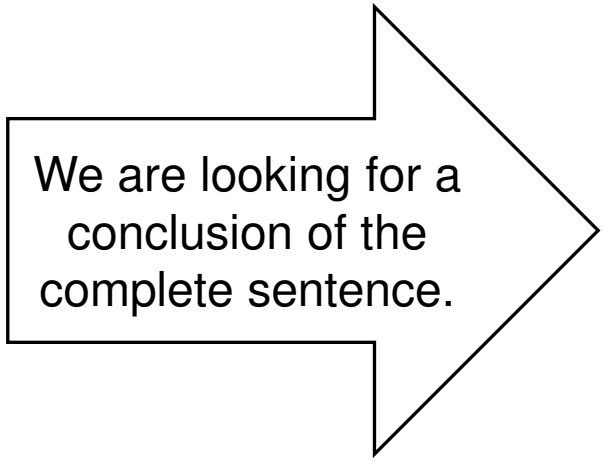
Sentiment analysis uses natural language processing to interpret and classify emotions in subjective data. Sentiment analysis is often used in business to detect sentiment in social data, gauge brand reputation, and understand customers.

User: It‘s a wonderful day.

- ***Sentiment: Positive***

User: It‘s a boring day.

- ***Sentiment: Negative***



User: It‘s a wonderful day.

- ***Sentiment: Positive***

User: It‘s a boring day.

- ***Sentiment: Negative***

Sentimental Analysis (or opinion mining)

Recurrent Neuronal Networks

Sentiment analysis uses natural language processing to interpret and classify emotions in subjective data. Sentiment analysis is often used in business to detect sentiment in social data, gauge brand reputation, and understand customers.

User: It‘s a wonderful day.

- ***Sentiment: Positive***

User: It‘s a boring day.

- ***Sentiment: Negative***

$$\hat{s}_t = \underset{s \in \text{Tags}}{\operatorname{argmax}} P(s|w_t)$$

We are looking for a conclusion of the complete sentence.

User: It‘s a wonderful day.

- ***Sentiment: Positive***

User: It‘s a boring day.

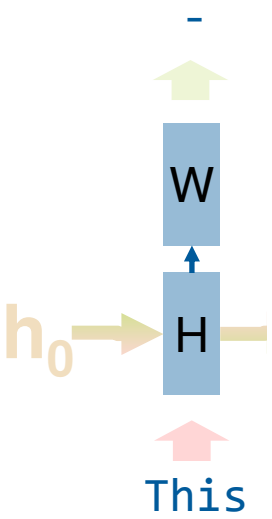
- ***Sentiment: Negative***

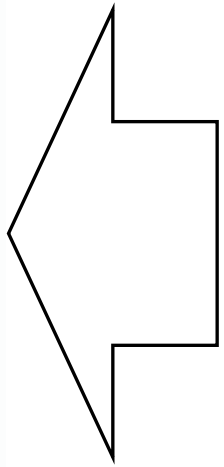
$$\hat{I} = \underset{I \in \text{Opinions}}{\operatorname{argmax}} P(I|\text{words})$$

Sentimental Analysis (or opinion mining)

Recurrent Neuronal Networks

Sentiment analysis uses natural language processing to interpret and classify emotions in subjective data. Sentiment analysis is often used in business to detect sentiment in social data, gauge brand reputation, and understand customers.





User: It's a wonderful day.

- ***Sentiment: Positive***

User: It's a boring day.

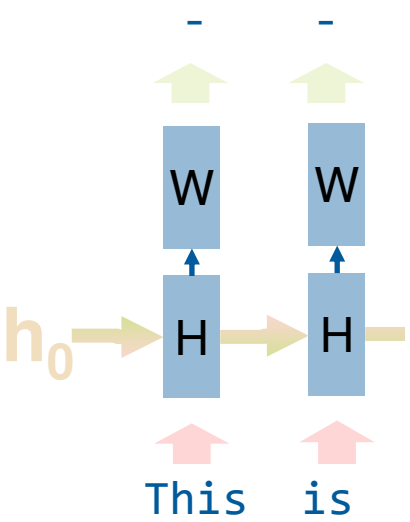
- ***Sentiment: Negative***

$$\hat{I} = \underset{I \in \text{Opinions}}{\operatorname{argmax}} P(I \mid \text{words})$$

Sentimental Analysis (or opinion mining)

Recurrent Neuronal Networks

Sentiment analysis uses natural language processing to interpret and classify emotions in subjective data. Sentiment analysis is often used in business to detect sentiment in social data, gauge brand reputation, and understand customers.



User: It's a wonderful day.

- ***Sentiment: Positive***

User: It's a boring day.

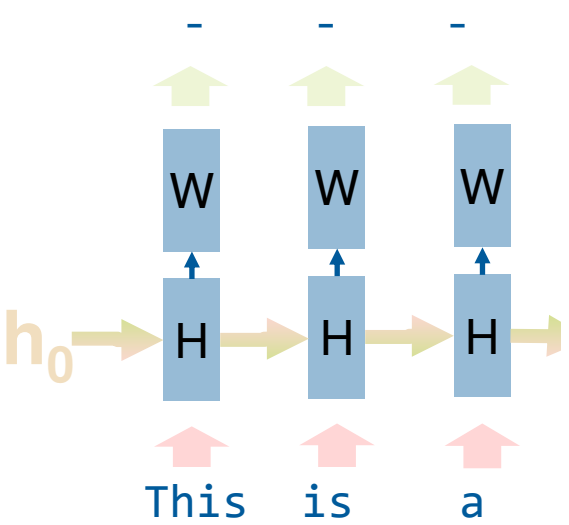
- ***Sentiment: Negative***

$$\hat{I} = \underset{I \in \text{Opinions}}{\text{argmax}}\, P(I \mid \text{words})$$

Sentimental Analysis (or opinion mining)

Recurrent Neuronal Networks

Sentiment analysis uses natural language processing to interpret and classify emotions in subjective data. Sentiment analysis is often used in business to detect sentiment in social data, gauge brand reputation, and understand customers.



User: It's a wonderful day.

- ***Sentiment: Positive***

User: It's a boring day.

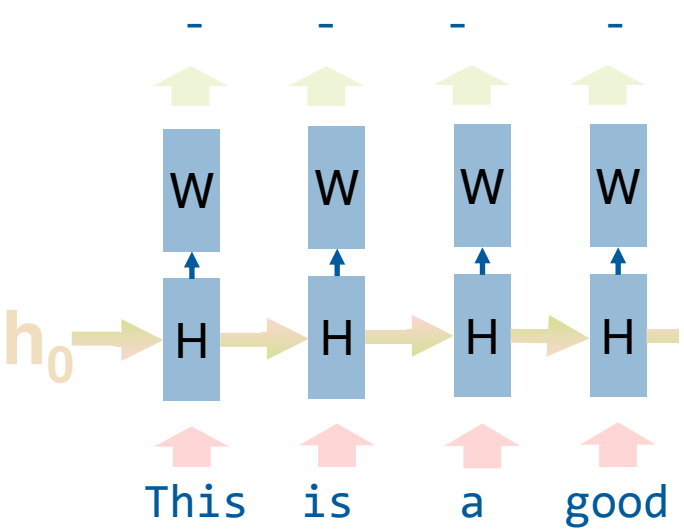
- ***Sentiment: Negative***

$$\hat{I} = \underset{I \in \text{Opinions}}{\text{argmax}}\, P(I \mid \text{words})$$

Sentimental Analysis (or opinion mining)

Recurrent Neuronal Networks

Sentiment analysis uses natural language processing to interpret and classify emotions in subjective data. Sentiment analysis is often used in business to detect sentiment in social data, gauge brand reputation, and understand customers.



User: It's a wonderful day.

- ***Sentiment: Positive***

User: It's a boring day.

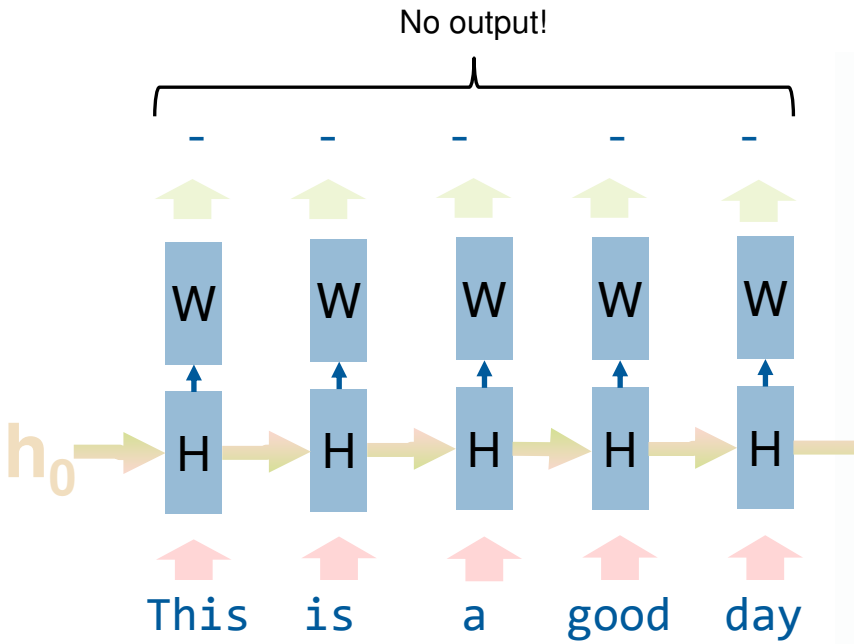
- ***Sentiment: Negative***

$$\hat{I} = \underset{I \in \text{Opinions}}{\text{argmax}}\, P(I \mid \text{words})$$

Sentimental Analysis (or opinion mining)

Recurrent Neuronal Networks

Sentiment analysis uses natural language processing to interpret and classify emotions in subjective data. Sentiment analysis is often used in business to detect sentiment in social data, gauge brand reputation, and understand customers.



User: It's a wonderful day.

- ***Sentiment: Positive***

User: It's a boring day.

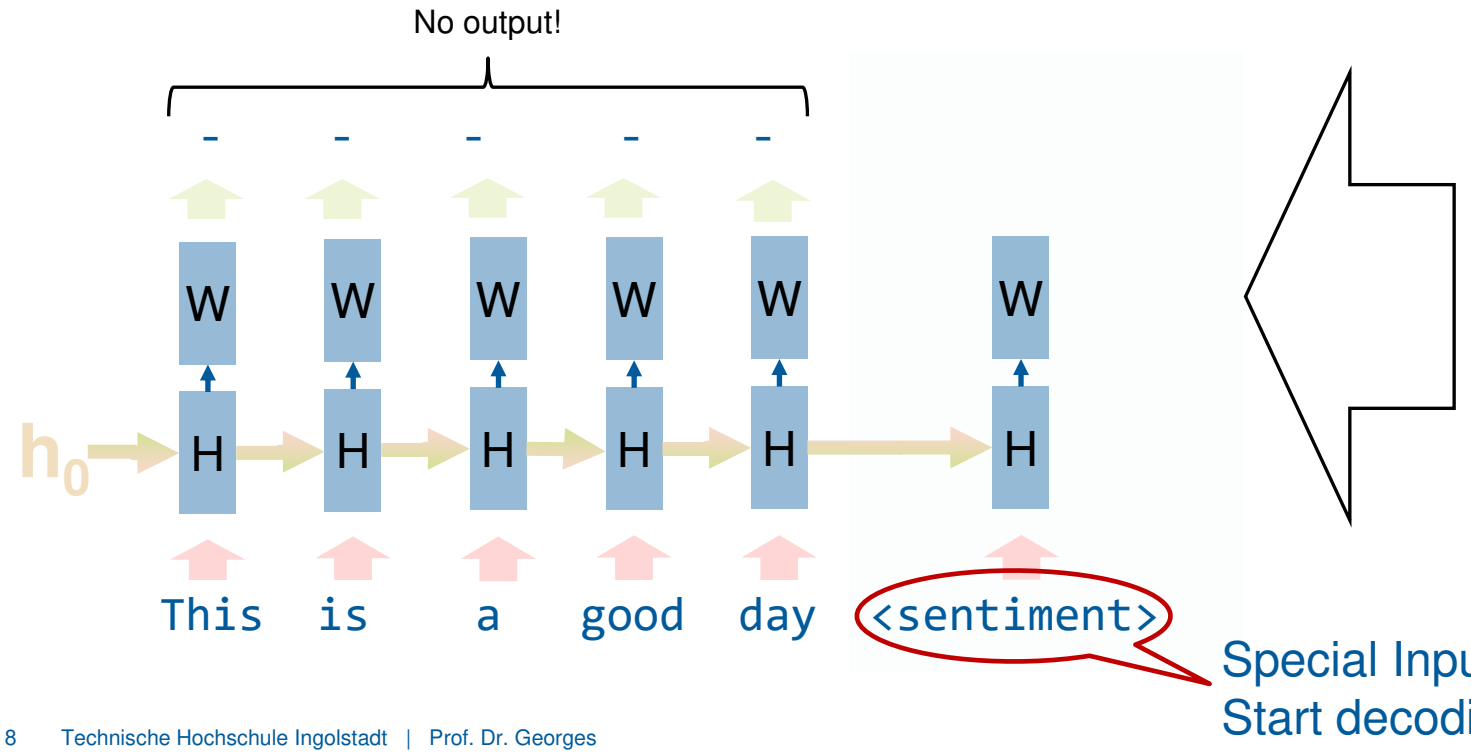
- ***Sentiment: Negative***

$$\hat{I} = \underset{I \in \text{Opinions}}{\text{argmax}}\, P(I \mid \text{words})$$

Sentimental Analysis (or opinion mining)

Recurrent Neuronal Networks

Sentiment analysis uses natural language processing to interpret and classify emotions in subjective data. Sentiment analysis is often used in business to detect sentiment in social data, gauge brand reputation, and understand customers.



User: It's a wonderful day.

- ***Sentiment: Positive***

User: It's a boring day.

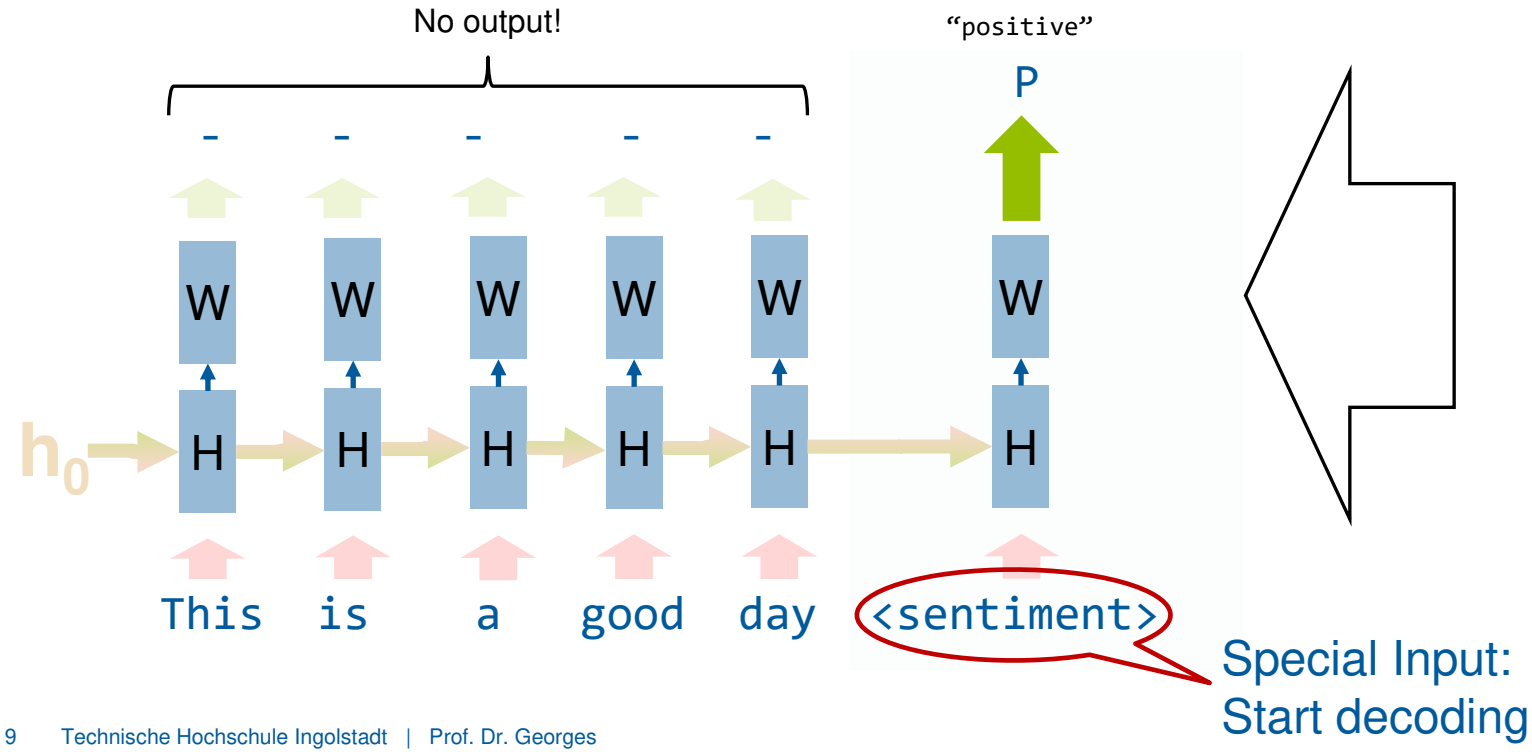
- ***Sentiment: Negative***

$$\hat{I} = \underset{I \in \text{Opinions}}{\operatorname{argmax}} P(I \mid \text{words})$$

Sentimental Analysis (or opinion mining)

Recurrent Neuronal Networks

Sentiment analysis uses natural language processing to interpret and classify emotions in subjective data. Sentiment analysis is often used in business to detect sentiment in social data, gauge brand reputation, and understand customers.



User: It's a wonderful day.

- ***Sentiment: Positive***

User: It's a boring day.

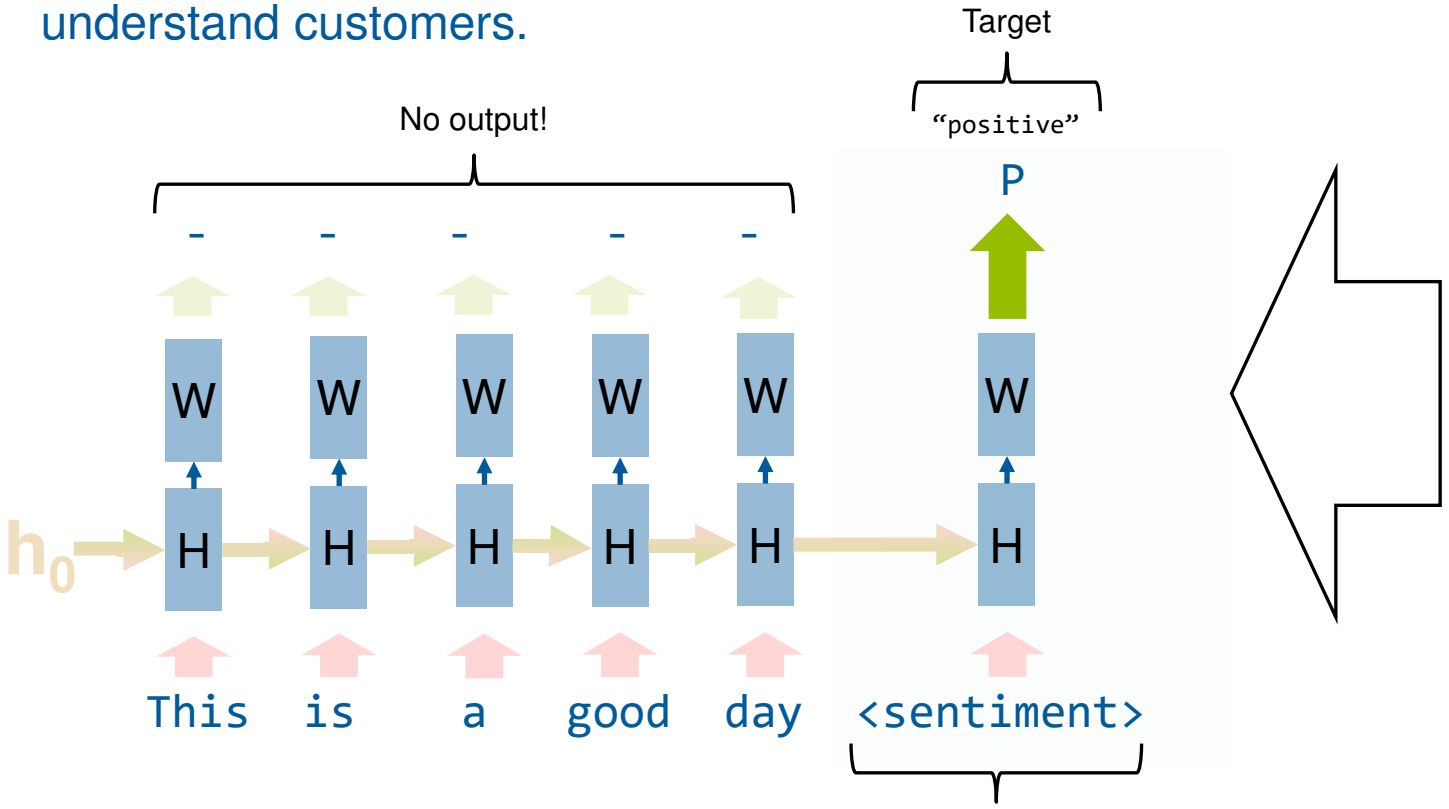
- ***Sentiment: Negative***

$$\hat{I} = \underset{I \in \text{Opinions}}{\operatorname{argmax}} P(I \mid \text{words})$$

Sentimental Analysis (or opinion mining)

Recurrent Neuronal Networks

Sentiment analysis uses natural language processing to interpret and classify emotions in subjective data. Sentiment analysis is often used in business to detect sentiment in social data, gauge brand reputation, and understand customers.



User: It's a wonderful day.

- ***Sentiment: Positive***

User: It's a boring day.

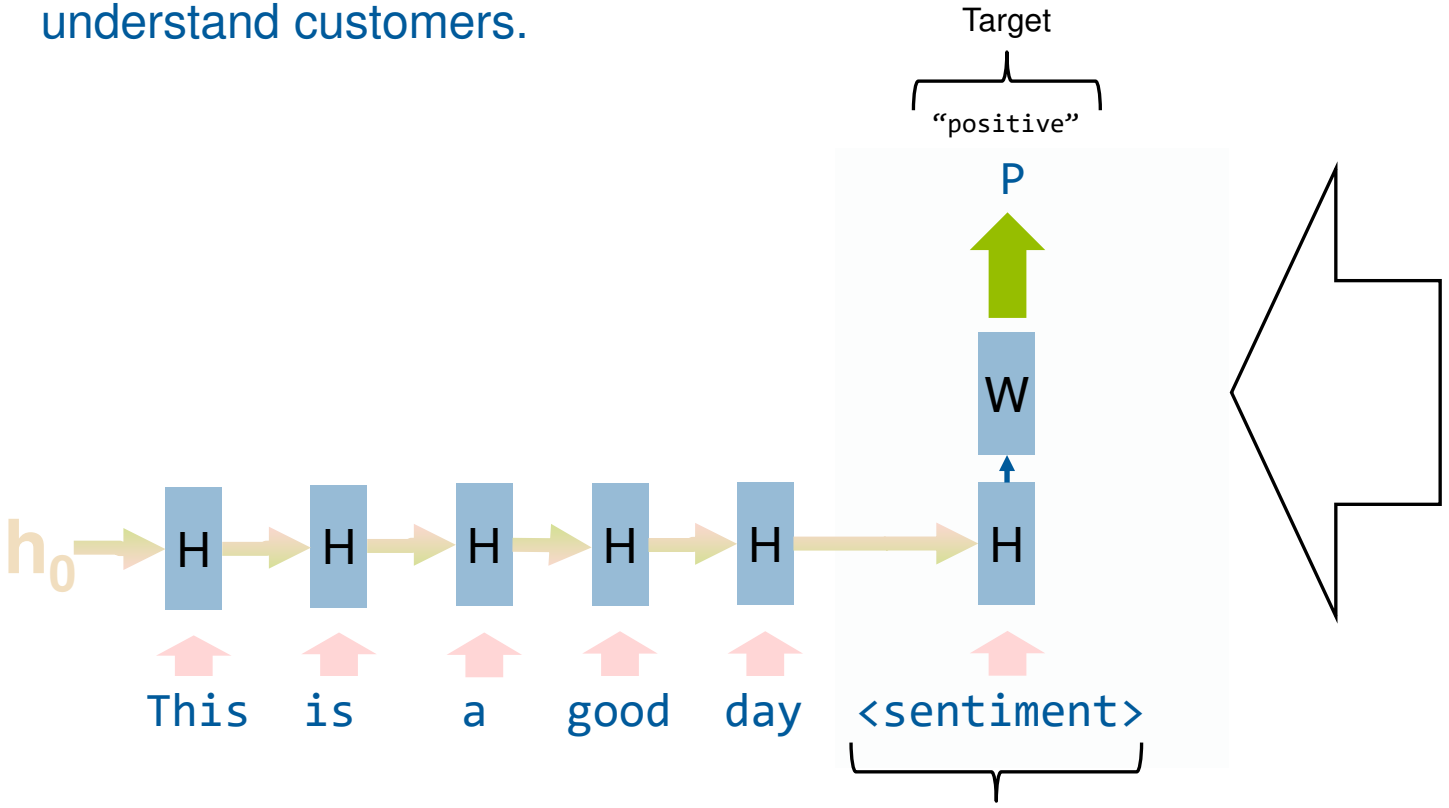
- ***Sentiment: Negative***

$$\hat{I} = \underset{I \in \text{Opinions}}{\text{argmax}}\, P(I \mid \text{words})$$

Sentimental Analysis (or opinion mining)

Recurrent Neuronal Networks

Sentiment analysis uses natural language processing to interpret and classify emotions in subjective data. Sentiment analysis is often used in business to detect sentiment in social data, gauge brand reputation, and understand customers.



User: It's a wonderful day.

- ***Sentiment: Positive***

User: It's a boring day.

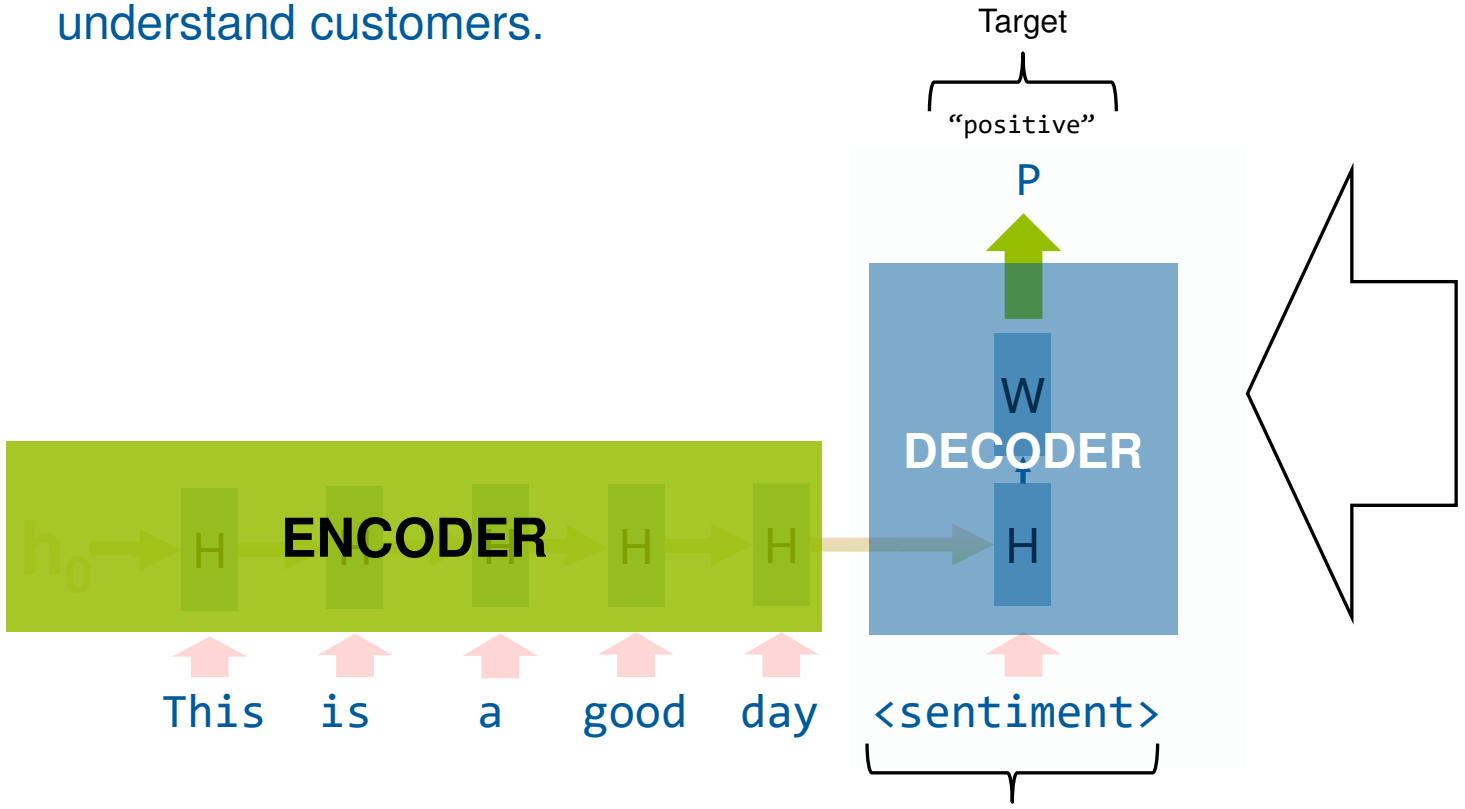
- ***Sentiment: Negative***

$$\hat{I} = \underset{I \in \text{Opinions}}{\text{argmax}}\, P(I \mid \text{words})$$

Sentimental Analysis (or opinion mining)

Recurrent Neuronal Networks

Sentiment analysis uses natural language processing to interpret and classify emotions in subjective data. Sentiment analysis is often used in business to detect sentiment in social data, gauge brand reputation, and understand customers.



User: It's a wonderful day.

- ***Sentiment: Positive***

User: It's a boring day.

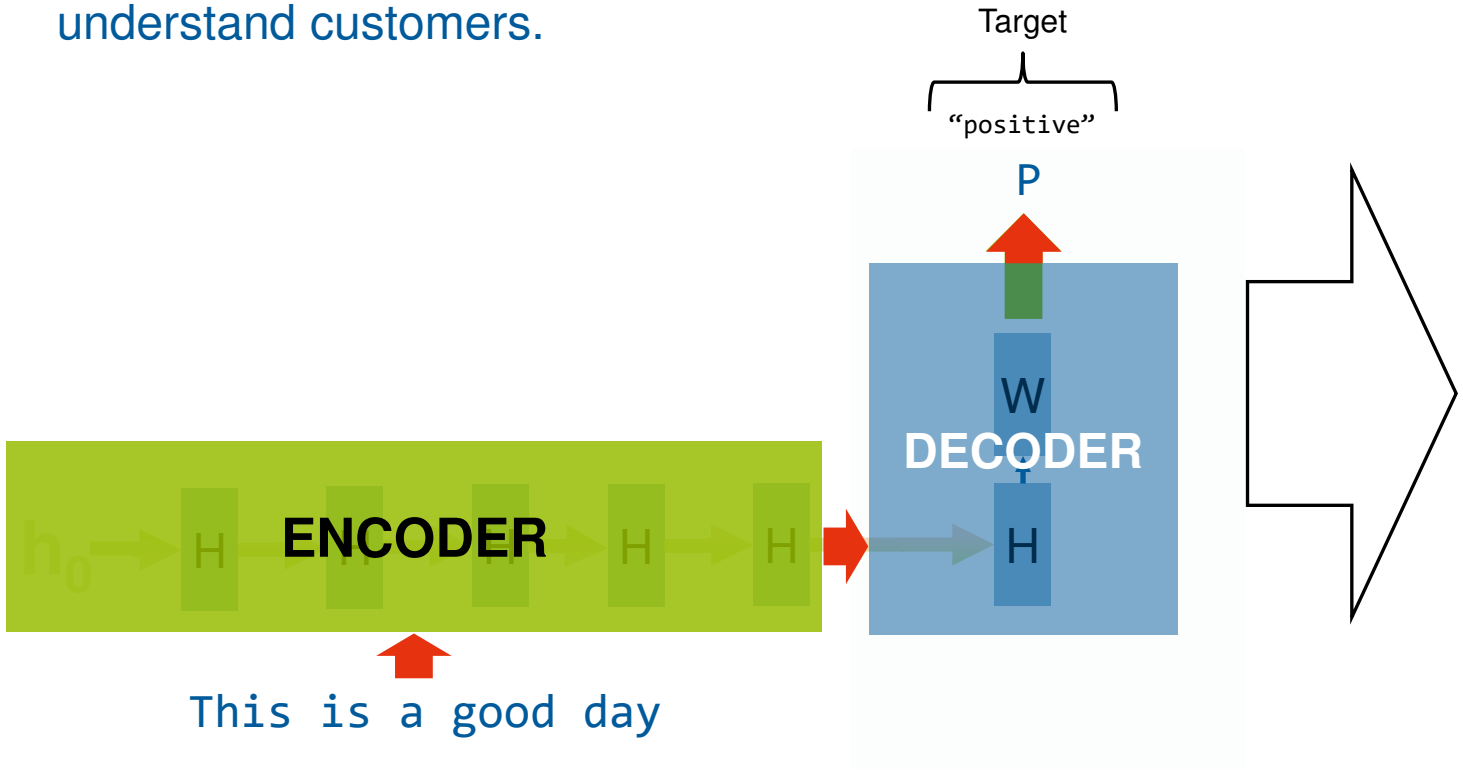
- ***Sentiment: Negative***

$$\hat{I} = \underset{I \in \text{Opinions}}{\operatorname{argmax}} P(I \mid \text{words})$$

Sentimental Analysis (or opinion mining)

Recurrent Neuronal Networks

Sentiment analysis uses natural language processing to interpret and classify emotions in subjective data. Sentiment analysis is often used in business to detect sentiment in social data, gauge brand reputation, and understand customers.



User: It's a wonderful day.

- ***Sentiment: Positive***

User: It's a boring day.

- ***Sentiment: Negative***

$$\hat{I} = \underset{I \in \text{Opinions}}{\text{argmax}}\ P(I|\text{words})$$

Intent detection

Definition

“Intent Detection is a vital component of any task-oriented conversational system. In order to understand the user’s current goal, the system must leverage its intent detector to classify the user’s utterance (provided in varied natural language) into one of several predefined classes, that is, intents.”

Domain	Intent	Utterance
Card	Lost	Could you assist me in finding my lost card?
Book	Restaurant	Book a table for one at a highly rated bistro
Play	Music	Play music by Damien Rice
Get	Weather	Weather conditions in East Pasadena

<Domain>.<Intent> e.g. Play.Music

Play is <Domain>

Music is <Intent>

https://arxiv.org/pdf/2003.04807.pdf

Intent detection

Definition

“Intent Detection is a vital component of any task-oriented conversational system. In order to understand the user’s current goal, the system must leverage its intent detector to classify the user’s utterance (provided in varied natural language) into one of several predefined classes, that is, intents.”

Example:

User: ***IT IS TO DARK HERE***

https://arxiv.org/pdf/2003.04807.pdf

Intent detection

Definition

“Intent Detection is a vital component of any task-oriented conversational system. In order to understand the user’s current goal, the system must leverage its intent detector to classify the user’s utterance (provided in varied natural language) into one of several predefined classes, that is, intents.”

Example:

User: ***IT IS TO DARK HERE***

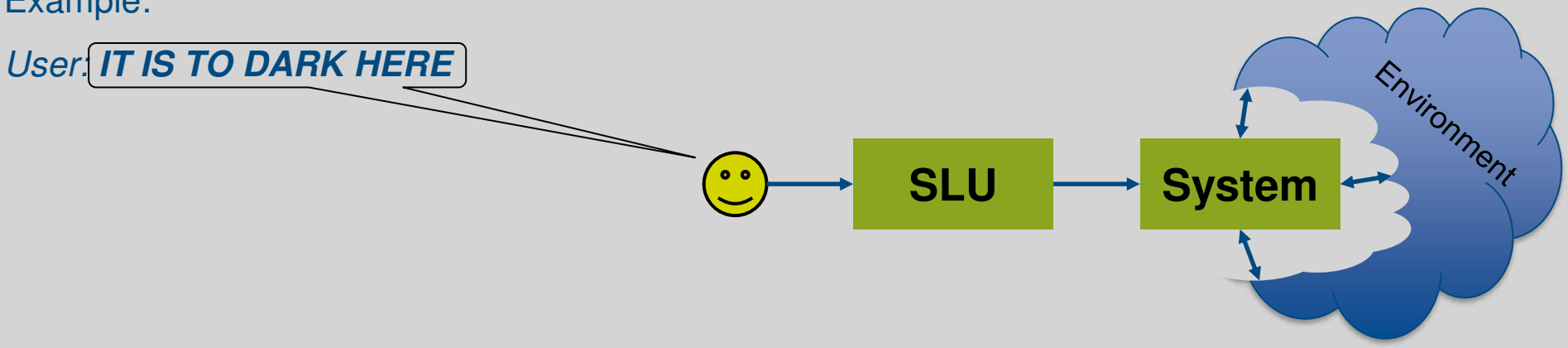
https://arxiv.org/pdf/2003.04807.pdf

Intent detection

Definition

“Intent Detection is a vital component of any task-oriented conversational system. In order to understand the user’s current goal, the system must leverage its intent detector to classify the user’s utterance (provided in varied natural language) into one of several predefined classes, that is, intents.”

Example:



https://arxiv.org/pdf/2003.04807.pdf

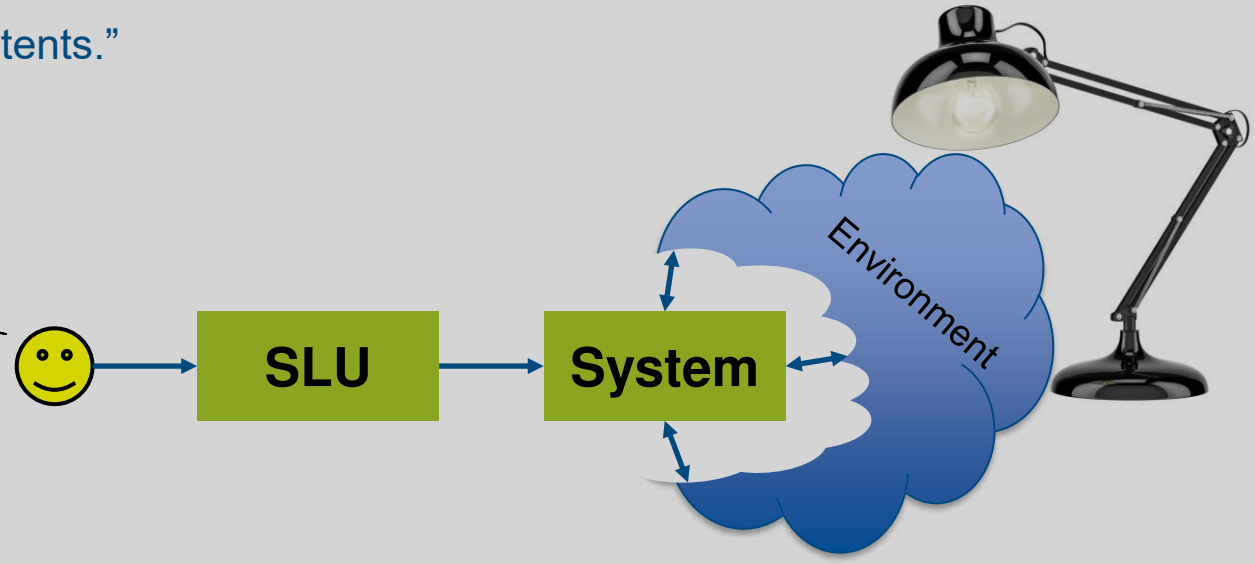
Intent detection

Definition

“Intent Detection is a vital component of any task-oriented conversational system. In order to understand the user’s current goal, the system must leverage its intent detector to classify the user’s utterance (provided in varied natural language) into one of several predefined classes, that is, intents.”

Example:

User: ***IT IS TO DARK HERE***



https://arxiv.org/pdf/2003.04807.pdf

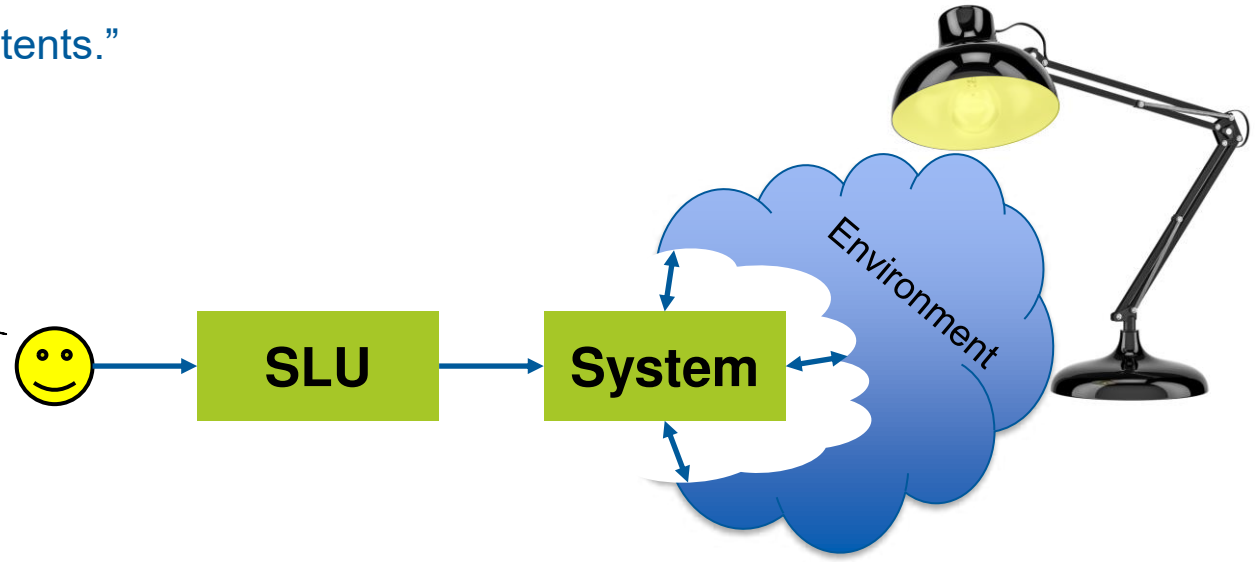
Intent detection

Definition

"Intent Detection is a vital component of any task-oriented conversational system. In order to understand the user's current goal, the system must leverage its intent detector to classify the user's utterance (provided in varied natural language) into one of several predefined classes, that is, intents."

Example:

User: ***IT IS TO DARK HERE***



https://arxiv.org/pdf/2003.04807.pdf

Intent detection

Definition

"Intent Detection is a vital component of any task-oriented conversational system. In order to understand the user's current goal, the system must leverage its intent detector to classify the user's utterance (provided in varied natural language) into one of several predefined classes, that is, intents."

Example:

User: ***IT IS TO DARK HERE***

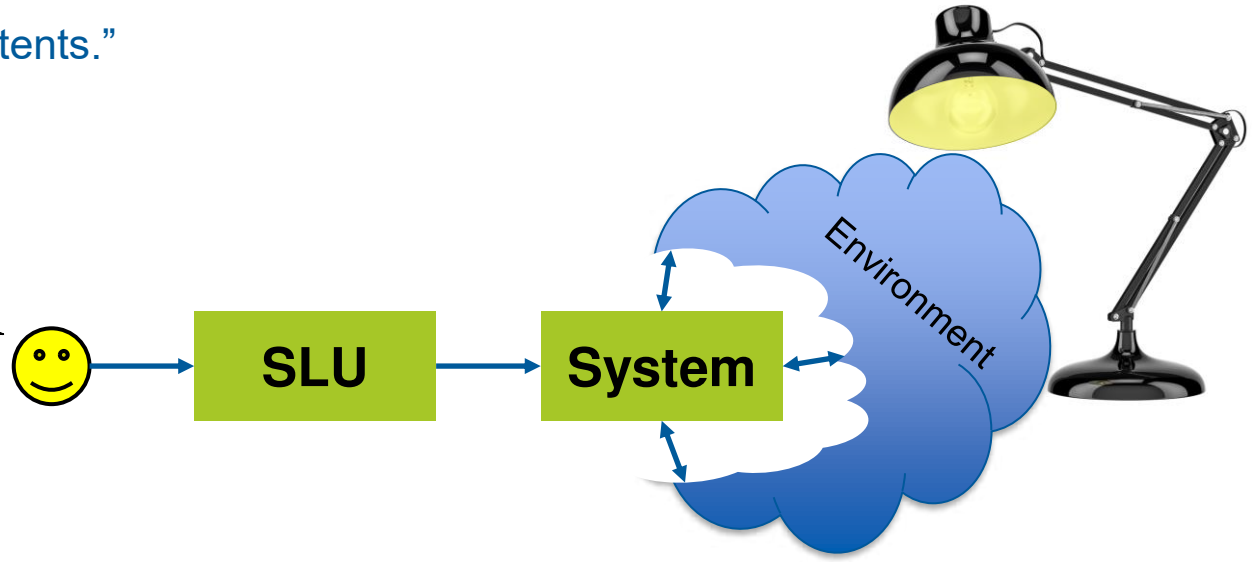
System: ***<turn light on>***

User: ***PLAY SOME MUSIC***

System: ***<turn radio on>***

User: ***TURN LEFT, PLEASE***

System: ***<turning the „car" left>***

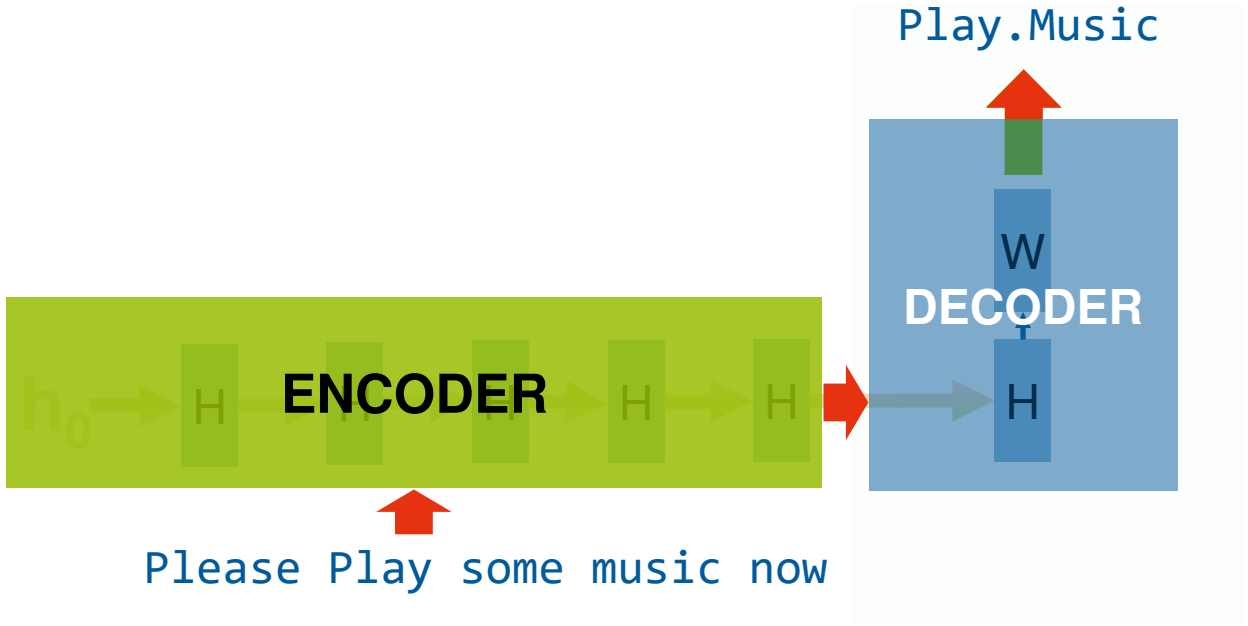


https://arxiv.org/pdf/2003.04807.pdf

Intent detection

Encoder-Decoder Architecture

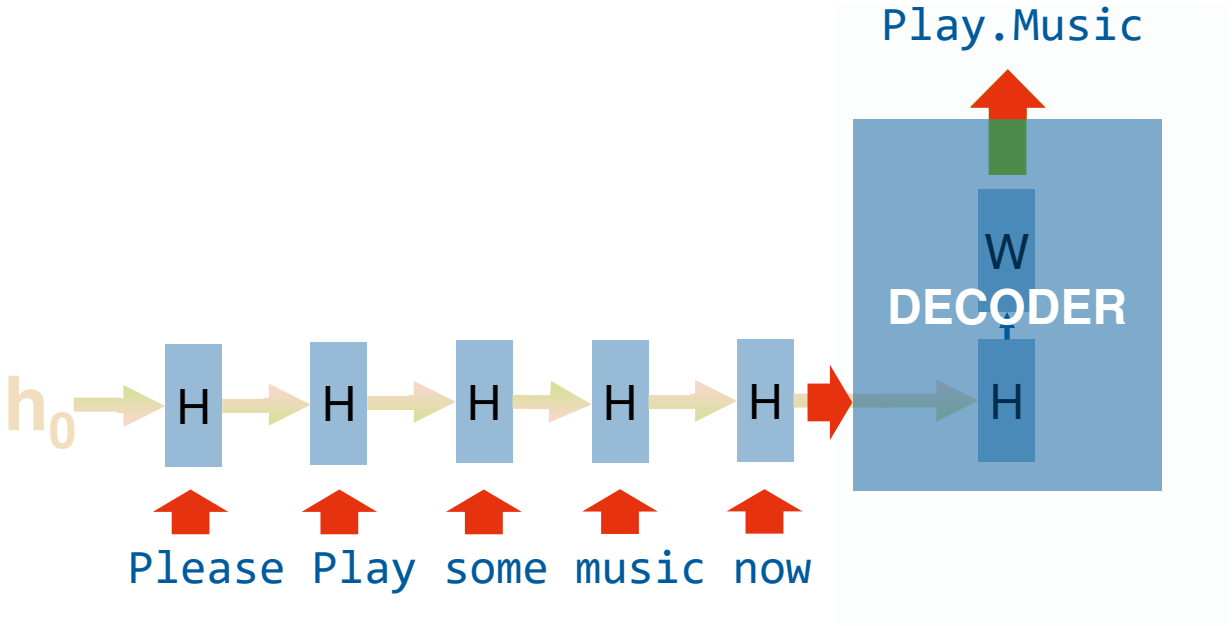
"In order to understand the user's current goal, the system must leverage its intent detector to classify the user's utterance into one of several predefined classes, that is, intents."



Intent detection

Encoder with Recurrent Neuronal Networks

"In order to understand the user's current goal, the system must leverage its intent detector to classify the user's utterance into one of several predefined classes, that is, intents."



Encoder:

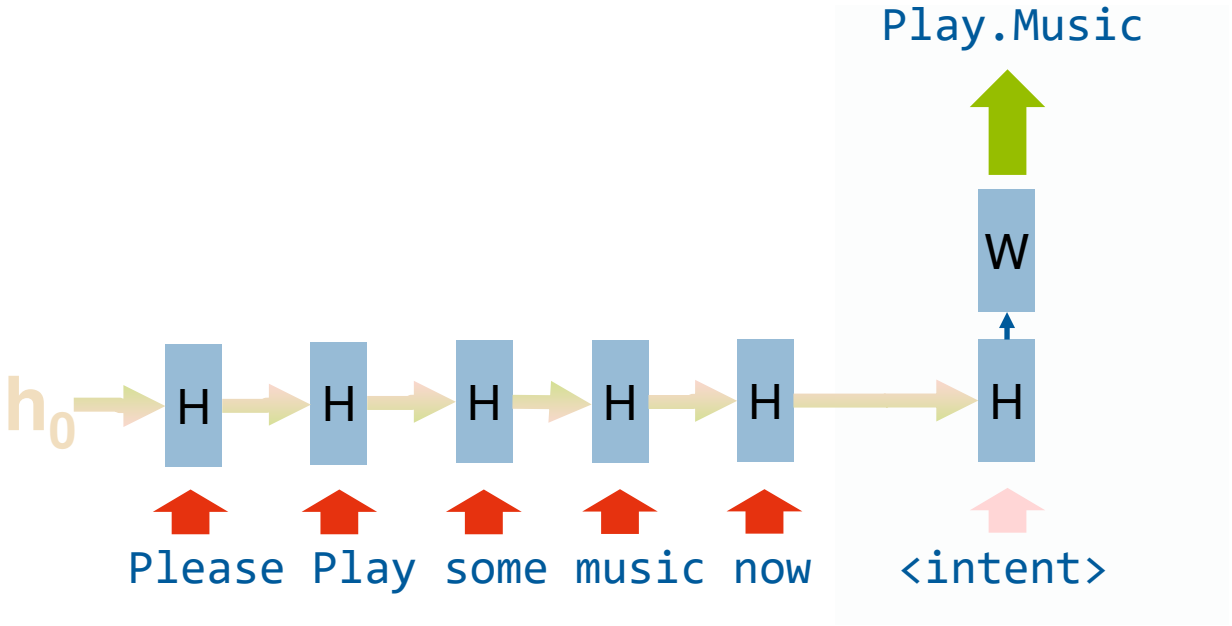
Word Embedding (input)

RNN, CNN, …

Intent detection

Decoder with Recurrent Neuronal Networks

“In order to understand the user’s current goal, the system must leverage its intent detector to classify the user’s utterance into one of several predefined classes, that is, intents.”



Encoder:

Word Embedding (input)

RNN, CNN, …

Decoder:

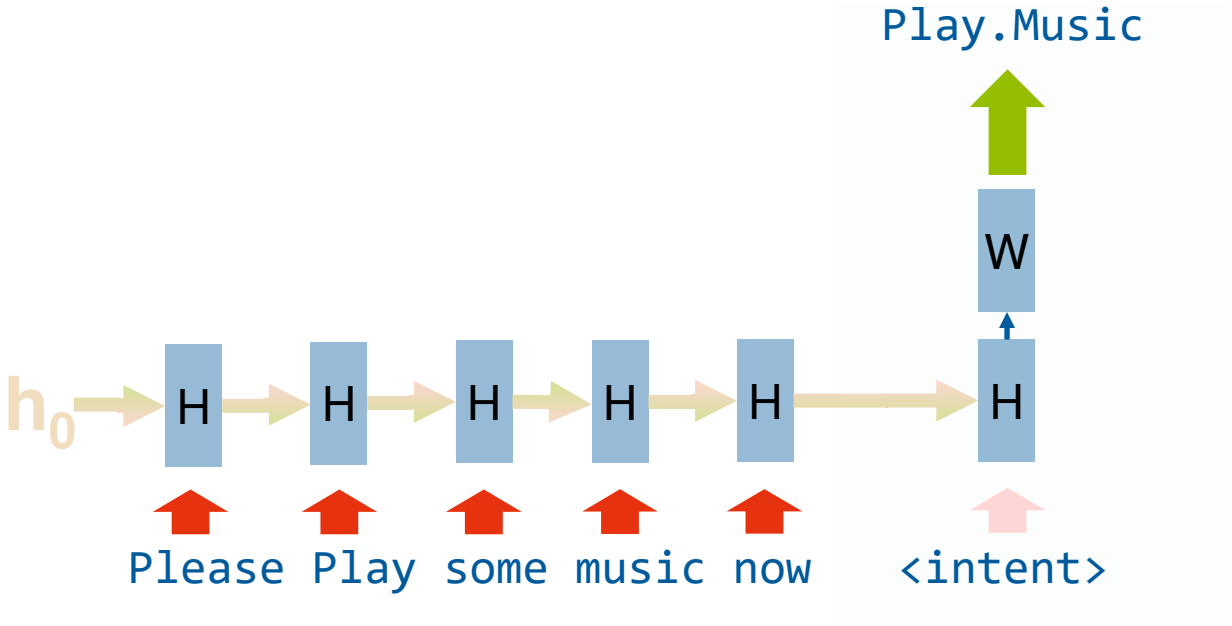
Word Embedding (output) with softmax

FFN, RNN, …

Intent detection

Encoder-Decoder with Recurrent Neuronal Networks

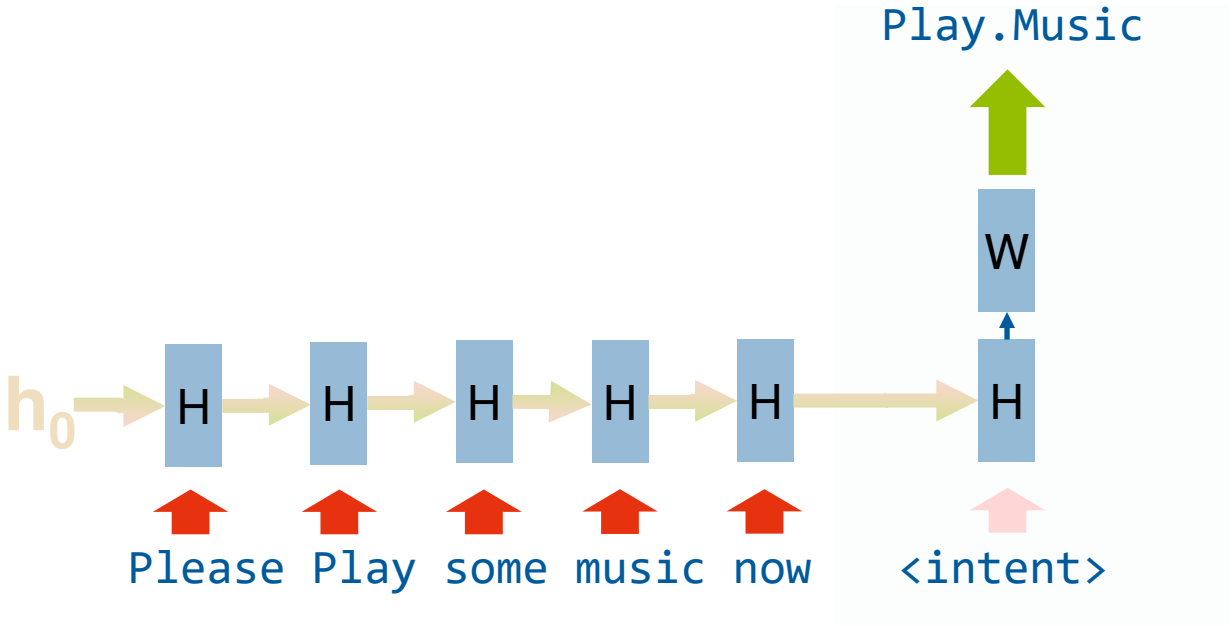
"In order to understand the user's current goal, the system must leverage its intent detector to classify the user's utterance into one of several predefined classes, that is, intents."



Intent detection

Encoder-Decoder with Recurrent Neuronal Networks

"In order to understand the user's current goal, the system must leverage its intent detector to classify the user's utterance into one of several predefined classes, that is, intents."



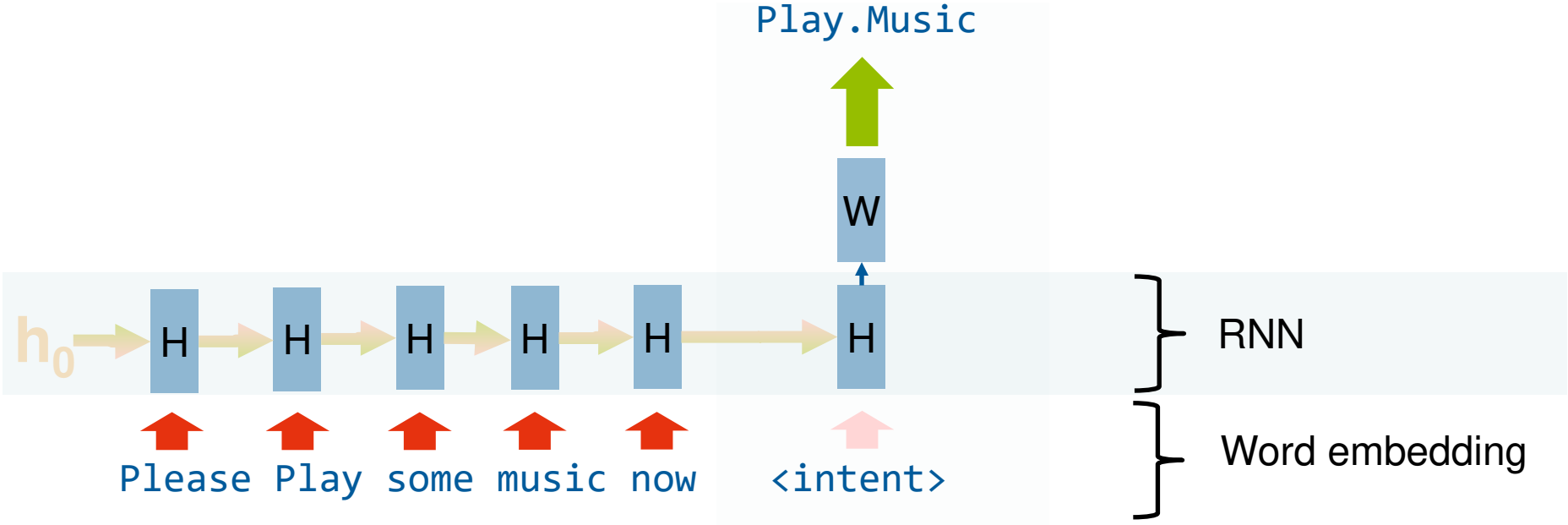
Word embedding

$$x_i = Ew_i$$

Intent detection

Encoder-Decoder with Recurrent Neuronal Networks

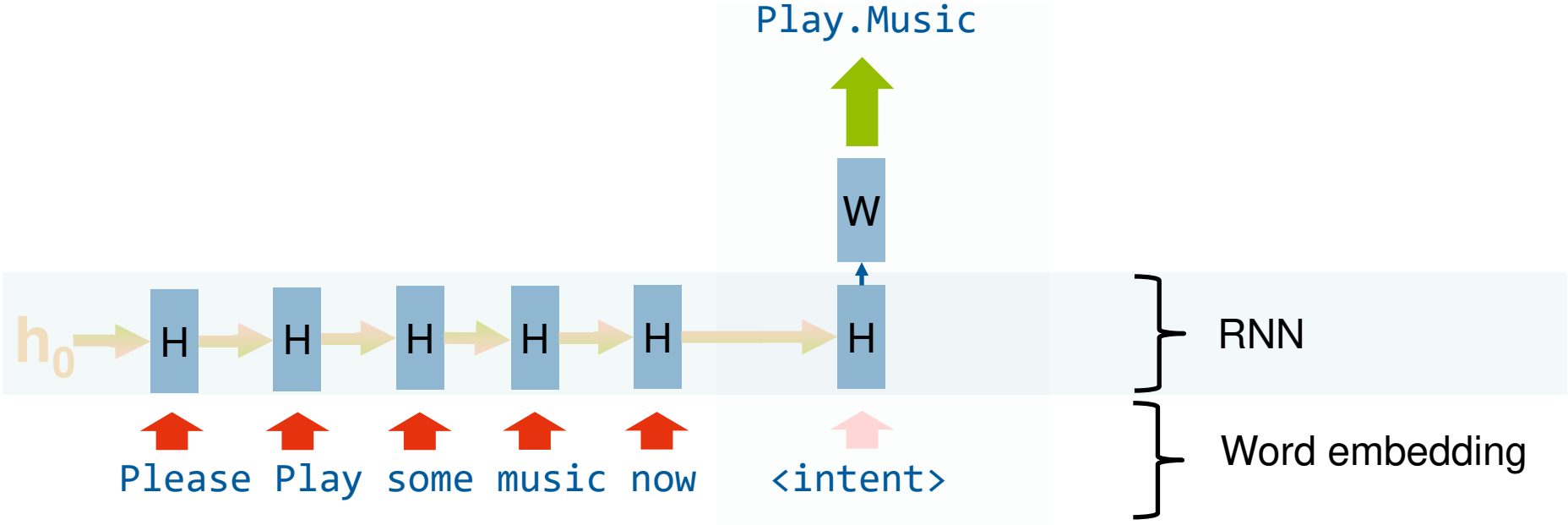
"In order to understand the user's current goal, the system must leverage its intent detector to classify the user's utterance into one of several predefined classes, that is, intents."



Intent detection

Encoder-Decoder with Recurrent Neuronal Networks

"In order to understand the user's current goal, the system must leverage its intent detector to classify the user's utterance into one of several predefined classes, that is, intents."



e.g. Elman RNN:

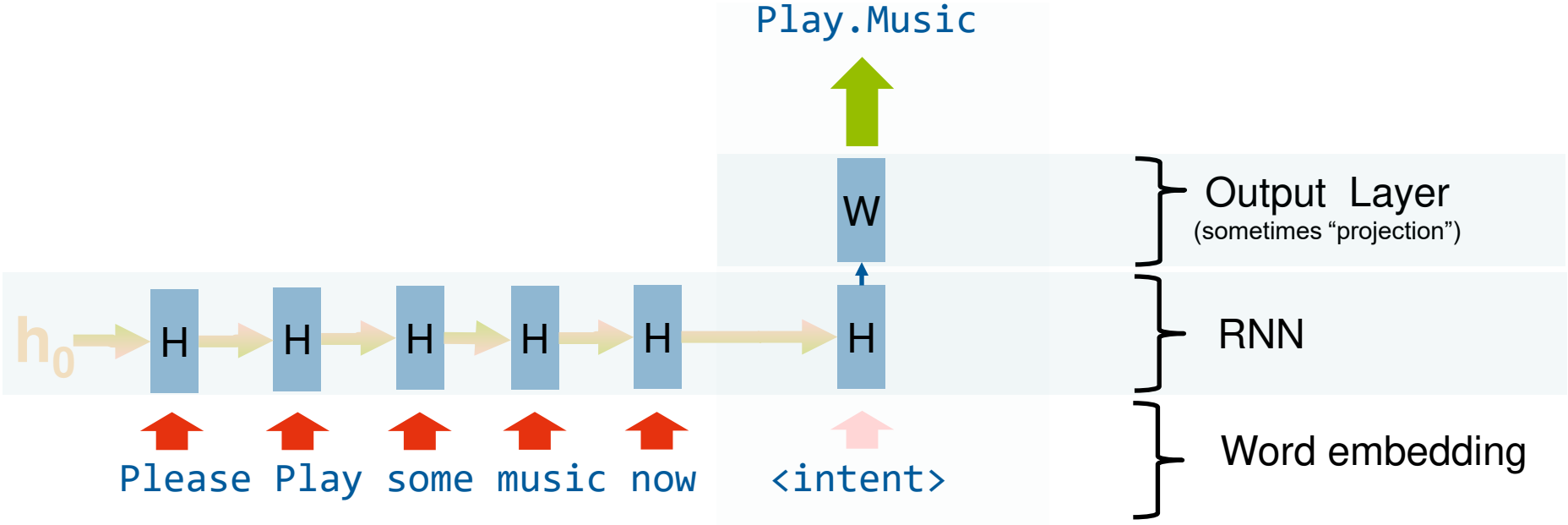
$$h_t = \sigma_h(W_h x_t + U_h h_{t-1})$$
$$y_t = \sigma_y(W_y h_t + b_y)$$

$$x_i = E w_i$$

Intent detection

Encoder-Decoder with Recurrent Neuronal Networks

"In order to understand the user's current goal, the system must leverage its intent detector to classify the user's utterance into one of several predefined classes, that is, intents."



What's the equation?

e.g. Elman RNN:

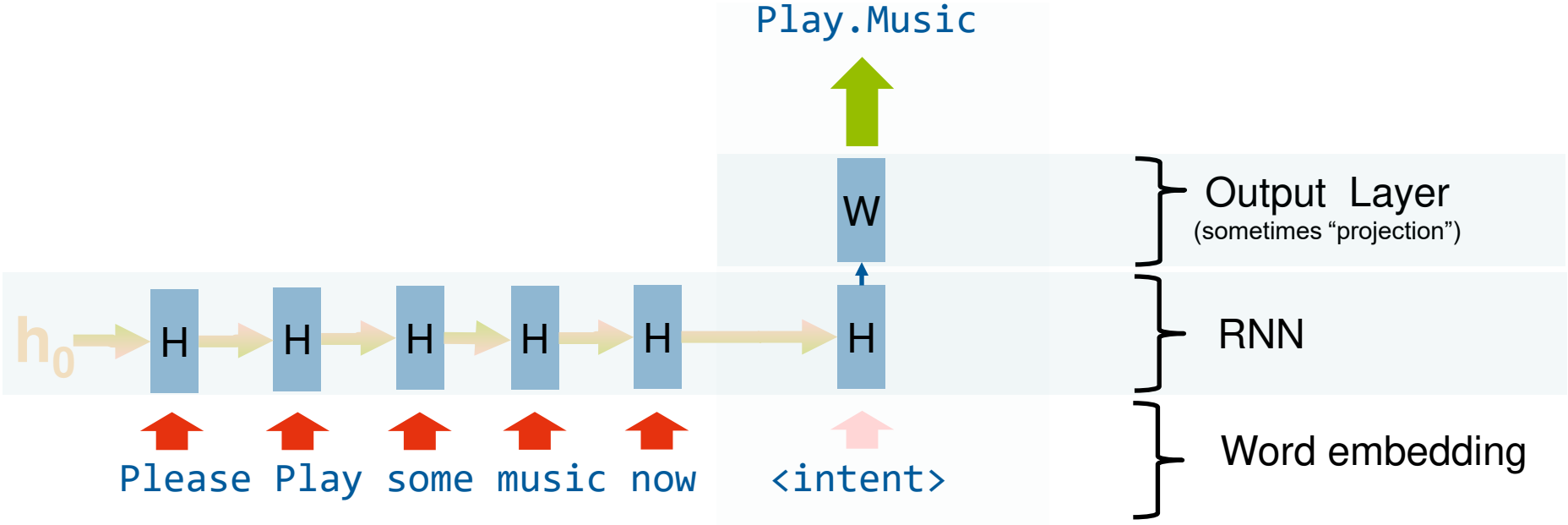
$$h_t = \sigma_h(W_h x_t + U_h h_{t-1} + b_h)$$
$$y_t = \sigma_y(W_y h_t + b_y)$$

$$x_i = E w_i$$

Intent detection

Encoder-Decoder with Recurrent Neuronal Networks

"In order to understand the user's current goal, the system must leverage its intent detector to classify the user's utterance into one of several predefined classes, that is, intents."



Whatever Neural Network

e.g. Elman RNN:

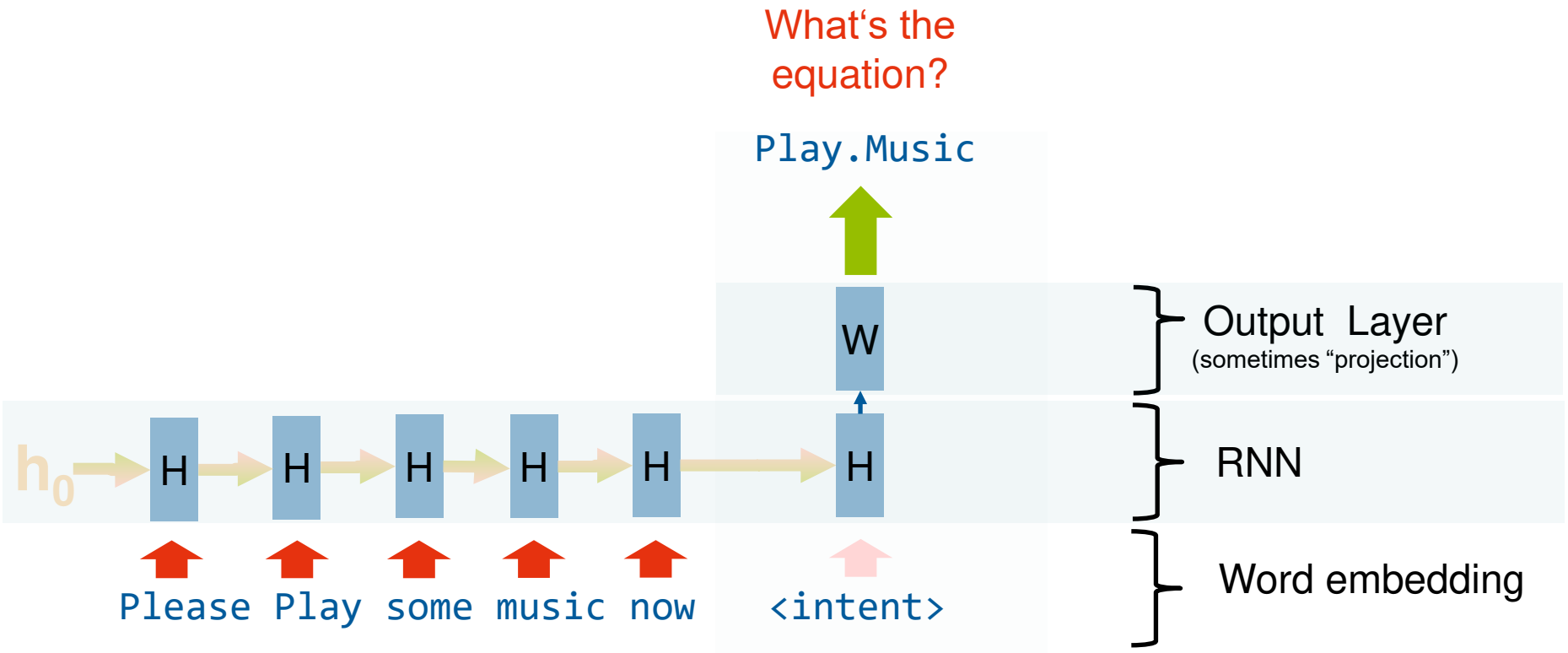
$$h_t = \sigma_h(W_h x_t + U_h h_{t-1} + b_h)$$
$$y_t = \sigma_y(W_y h_t + b_y)$$

$$x_i = E w_i$$

Intent detection

Encoder-Decoder with Recurrent Neuronal Networks

"In order to understand the user's current goal, the system must leverage its intent detector to classify the user's utterance into one of several predefined classes, that is, intents."



Whatever Neural Network

e.g. Elman RNN:

$$h_t = \sigma_h(W_h x_t + U_h h_{t-1} + b_h)$$
$$y_t = \sigma_y(W_y h_t + b_y)$$

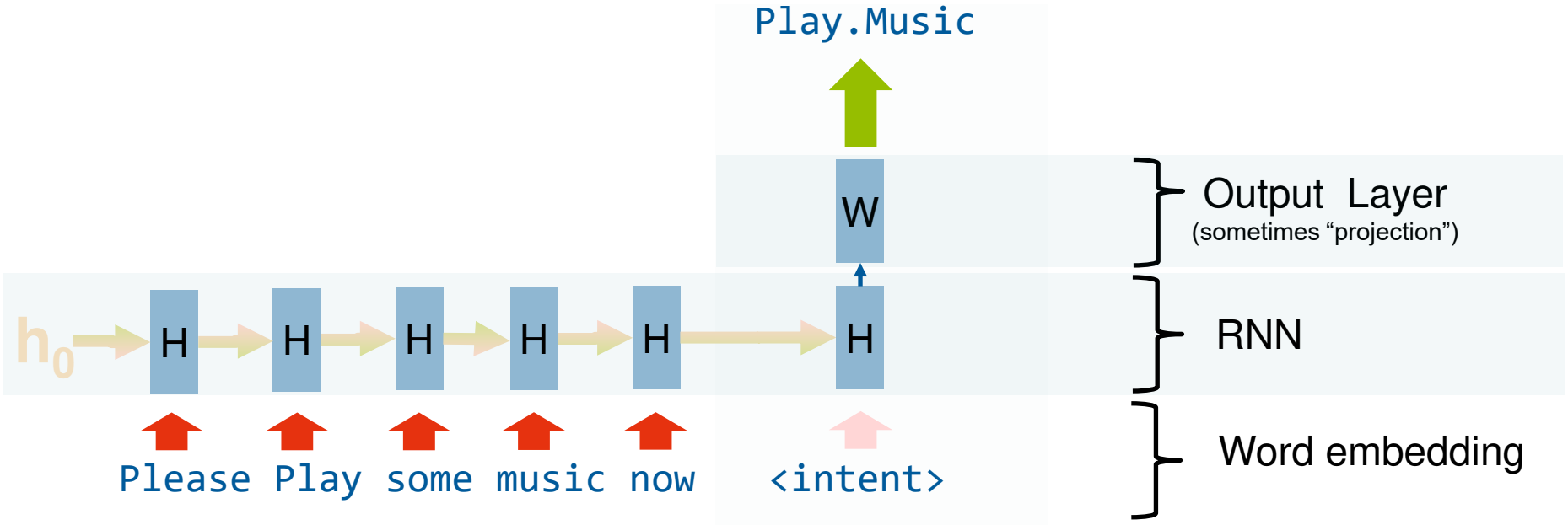
$$x_i = E w_i$$

Intent detection

Encoder-Decoder with Recurrent Neuronal Networks

"In order to understand the user's current goal, the system must leverage its intent detector to classify the user's utterance into one of several predefined classes, that is, intents."

$$\hat{I} = \underset{I \in \text{Intents}}{\text{argmax}}\, P(I|\text{words})$$



Whatever Neural Network

e.g. Elman RNN:

$$h_t = \sigma_h(W_h x_t + U_h h_{t-1} + b_h)$$
$$y_t = \sigma_y(W_y h_t + b_y)$$

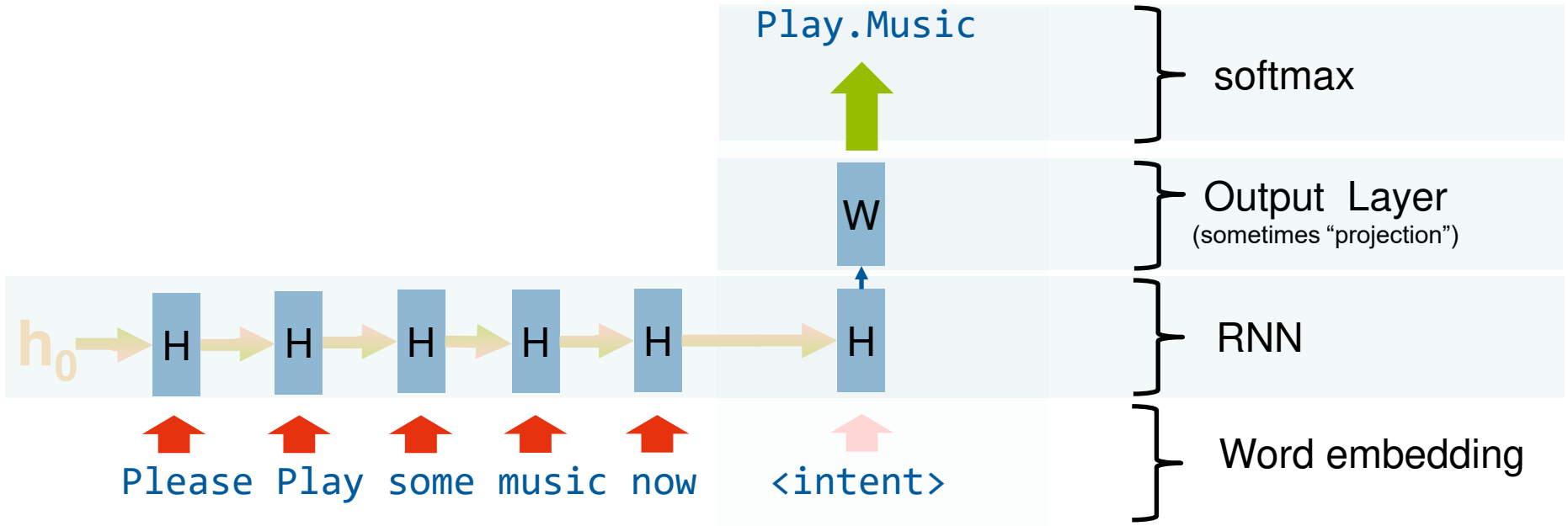
$$x_i = E w_i$$

Intent detection

Encoder-Decoder with Recurrent Neuronal Networks

"In order to understand the user's current goal, the system must leverage its intent detector to classify the user's utterance into one of several predefined classes, that is, intents."

$$\hat{I} = \underset{I \in \text{Intents}}{\text{argmax}}\, P(I|\text{words})$$



What's the equation?

Whatever Neural Network

e.g. Elman RNN:

$$h_t = \sigma_h(W_h x_t + U_h h_{t-1} + b_h)$$
$$y_t = \sigma_y(W_y h_t + b_y)$$

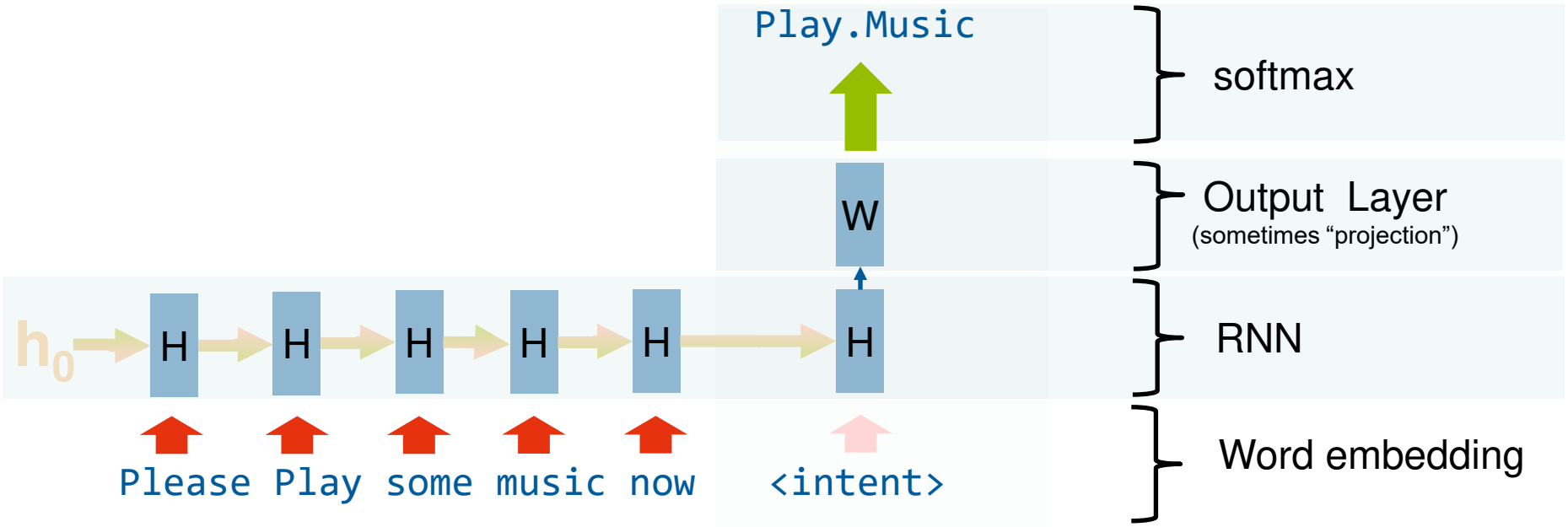
$$x_i = E w_i$$

Intent detection

Encoder-Decoder with Recurrent Neuronal Networks

"In order to understand the user's current goal, the system must leverage its intent detector to classify the user's utterance into one of several predefined classes, that is, intents."

$$\hat{I} = \underset{I \in \text{Intents}}{\text{argmax}}\, P(I|\text{words})$$



$$\sigma(\mathbf{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}}$$

Whatever Neural Network

e.g. Elman RNN:

$$h_t = \sigma_h(W_h x_t + U_h h_{t-1} + b_h)$$
$$y_t = \sigma_y(W_y h_t + b_y)$$

$$x_i = E w_i$$

Intent detection

Implementation with Bag-Of-Words and Feed Forward Network

“Intent Detection is a vital component of any task-oriented conversational system. In order to understand the user’s current goal, the system must leverage its intent detector to classify the user’s utterance (provided in varied natural language) into one of several predefined classes, that is, intents.”

Example:

Play some music → Play.Music

Let’s formalize it:
What is the equation?

https://arxiv.org/pdf/2003.04807.pdf

Intent detection

Implementation with Bag-Of-Words and Feed Forward Network

“Intent Detection is a vital component of any task-oriented conversational system. In order to understand the user’s current goal, the system must leverage its intent detector to classify the user’s utterance (provided in varied natural language) into one of several predefined classes, that is, intents.”

Example:

Play some music

Play.Music

$$\hat{I} = \underset{I \in \text{Intents}}{\operatorname{argmax}} P(I|\text{words})$$

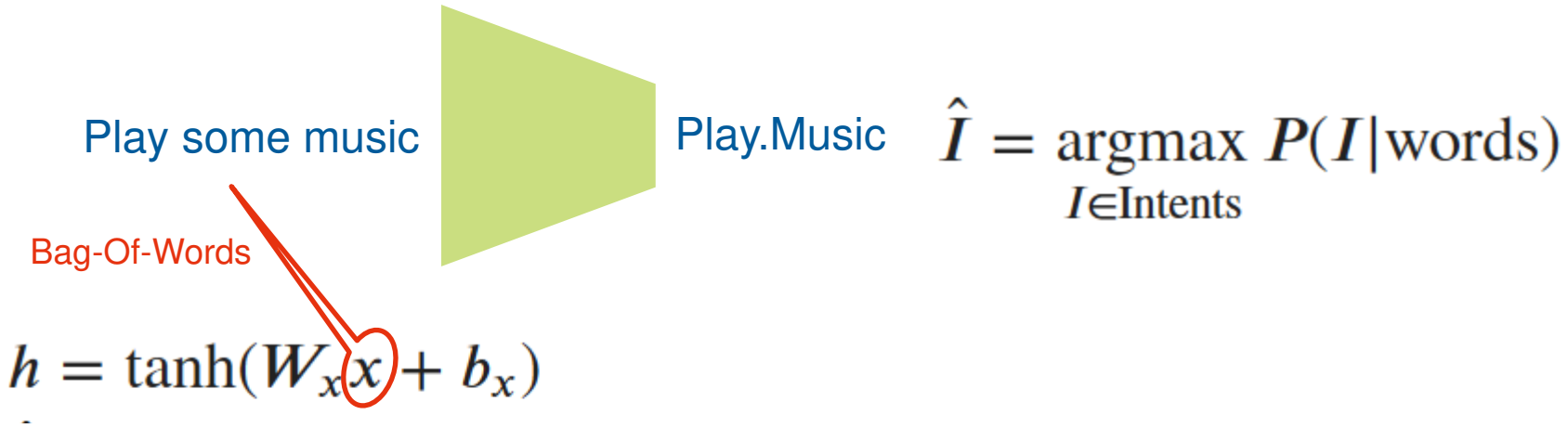
https://arxiv.org/pdf/2003.04807.pdf

Intent detection

Implementation with Bag-Of-Words and Feed Forward Network

"Intent Detection is a vital component of any task-oriented conversational system. In order to understand the user's current goal, the system must leverage its intent detector to classify the user's utterance (provided in varied natural language) into one of several predefined classes, that is, intents."

Example:



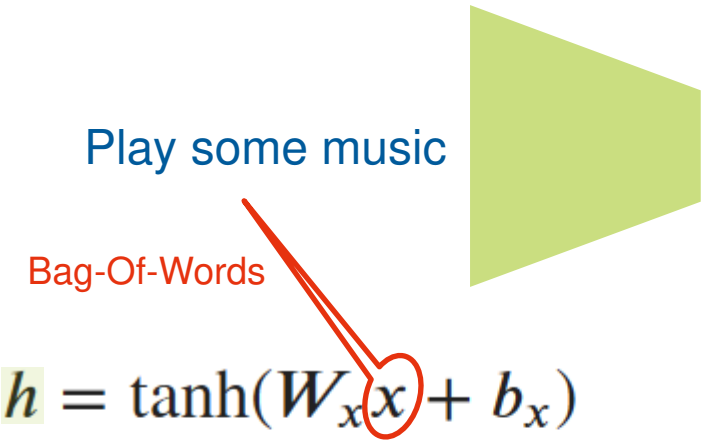
https://arxiv.org/pdf/2003.04807.pdf

Intent detection

Implementation with Bag-Of-Words and Feed Forward Network

"Intent Detection is a vital component of any task-oriented conversational system. In order to understand the user's current goal, the system must leverage its intent detector to classify the user's utterance (provided in varied natural language) into one of several predefined classes, that is, intents."

Example:



Play.Music

$$\hat{I} = \underset{I \in \text{Intents}}{\text{argmax}}\, P(I|\text{words})$$

$$h = \tanh(W_x x + b_x)$$

$$\hat{I} = \underset{I \in \text{Intents}}{\text{argmax}}(\text{softmax}(W_h h + b_h)_I)$$

https://arxiv.org/pdf/2003.04807.pdf

Neural Machine Translation

Definition

An approach to machine translation that uses an artificial neural network to predict the likelihood of a sequence of words, typically modeling entire sentences in a single integrated model.

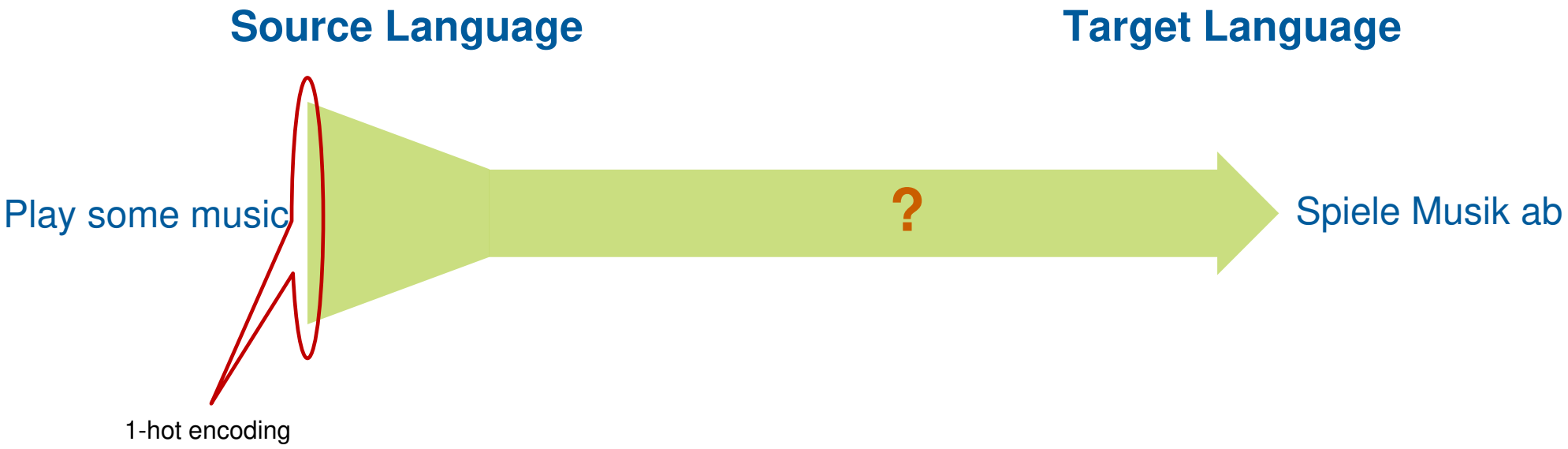
Source Language **Target Language**

https://arxiv.org/pdf/2003.04807.pdf

Neural Machine Translation

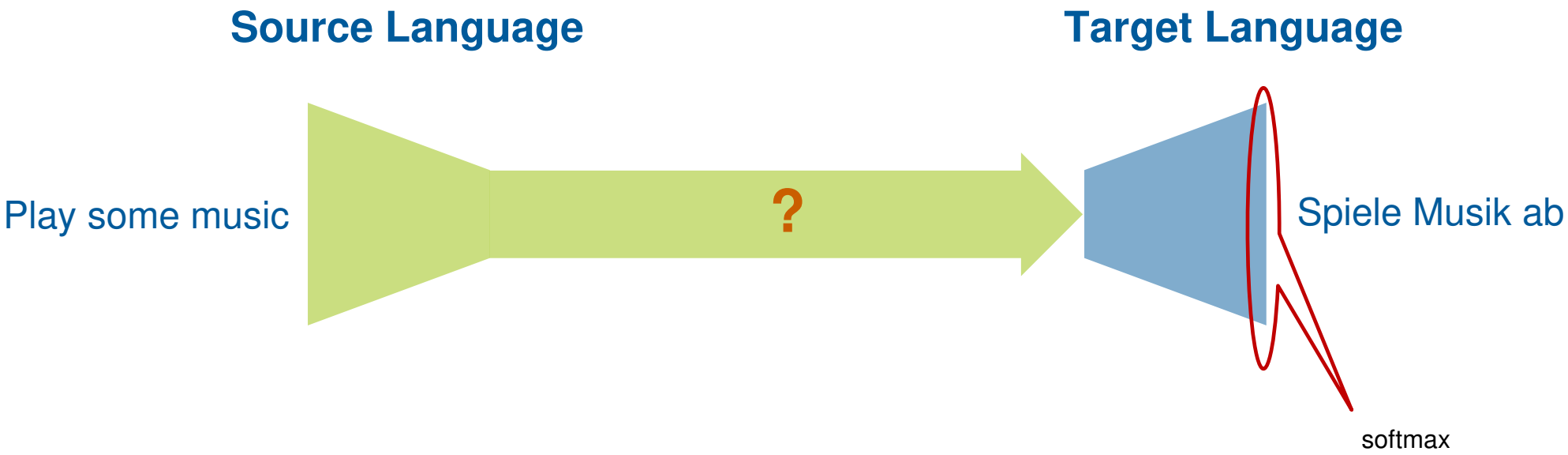
Definition

An approach to machine translation that uses an artificial neural network to predict the likelihood of a sequence of words, typically modeling entire sentences in a single integrated model.



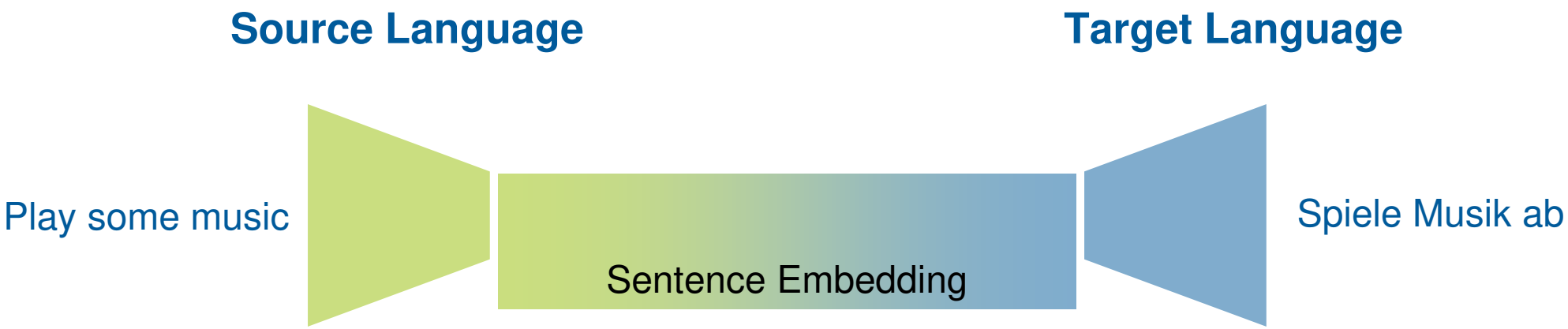
https://arxiv.org/pdf/2003.04807.pdf

An approach to machine translation that uses an artificial neural network to predict the likelihood of a sequence of words, typically modeling entire sentences in a single integrated model.



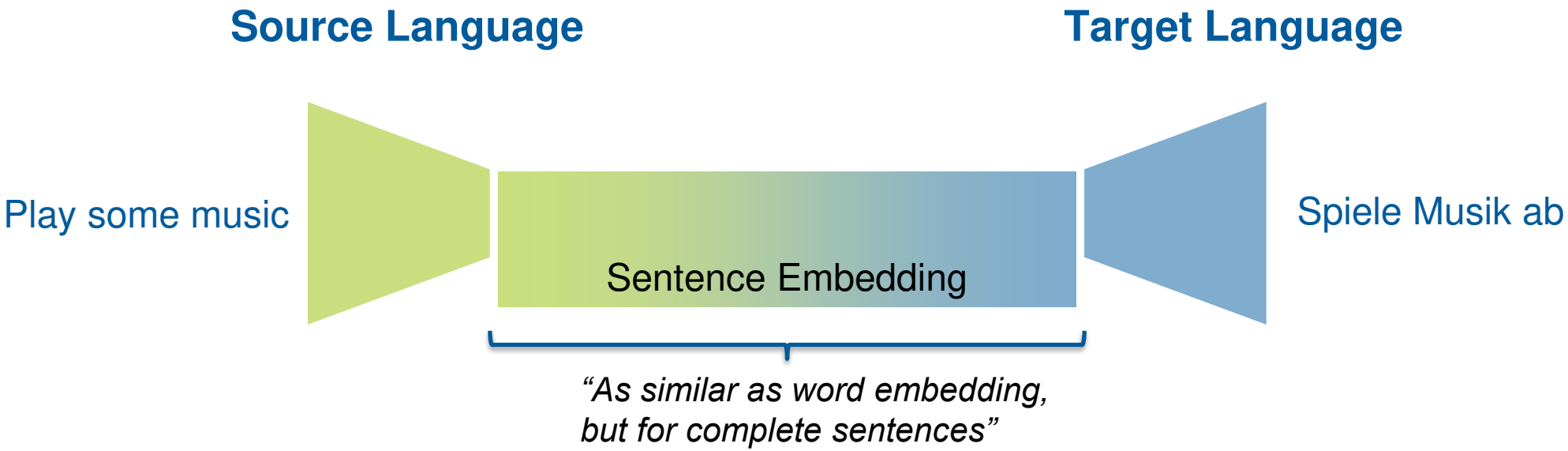
https://arxiv.org/pdf/2003.04807.pdf

An approach to machine translation that uses an artificial neural network to predict the likelihood of a sequence of words, typically modeling entire sentences in a single integrated model.



https://arxiv.org/pdf/2003.04807.pdf

An approach to machine translation that uses an artificial neural network to predict the likelihood of a sequence of words, typically modeling entire sentences in a single integrated model.

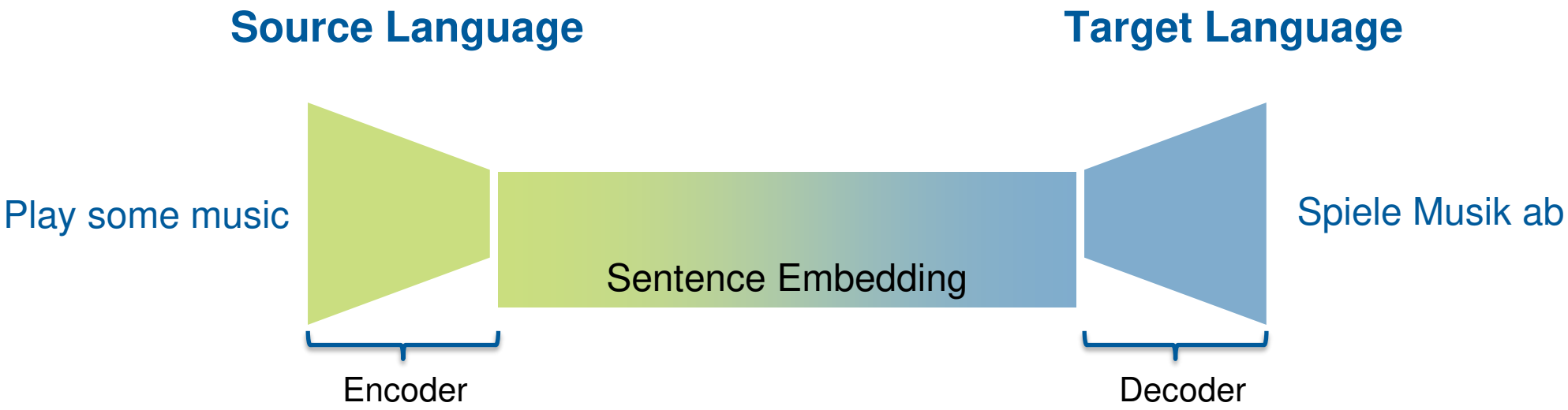


https://arxiv.org/pdf/2003.04807.pdf

Neural Machine Translation

Definition

An approach to machine translation that uses an artificial neural network to predict the likelihood of a sequence of words, typically modeling entire sentences in a single integrated model.



https://arxiv.org/pdf/2003.04807.pdf

Predict next word y_t, given …

- previous words $\{y_i\}_{i<t}$

$$P(y_t | \{y_i\}_{i<t})$$

Predict next word y_t, given …

- previous words $\{y_i\}_{i<t}$
- additional input x

$$P(y_t | \{y_i\}_{i<t}, x)$$

Example: Neural Machine Translation (NMT)

and conditional Language Modelling

Predict next word y_t, given …

- previous words $\{y_i\}_{i<t}$
- additional input x

$$P(y_t | \{y_i\}_{i<t}, x)$$

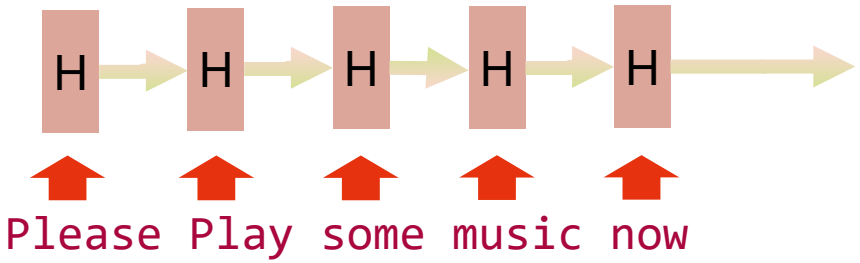
Neural Machine Translation

Source sequence: `Please Play some music now`

Target sequence: `Bitte Musik jetzt abspielen`

Example: NMT

Encoder-Decoder with Recurrent Neuronal Networks and Autoregression



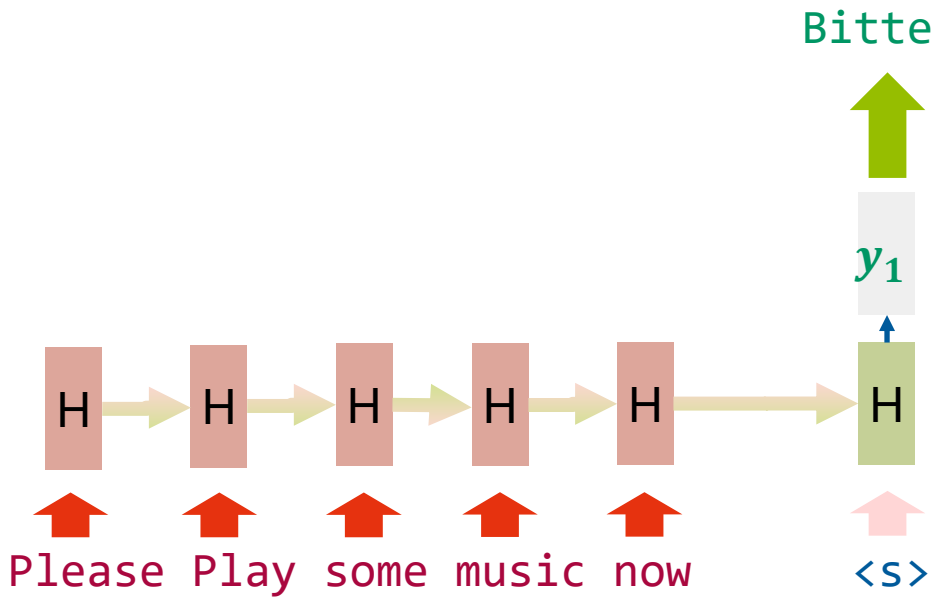
Example: NMT

Encoder-Decoder with Recurrent Neuronal Networks and Autoregression

Minimize $J = \sum_{t=1}^{T} -\log P(\boldsymbol{y_t}|\{y_i\}_{i<t}, x) = J_1$

↓

negative log prob of „Bitte“

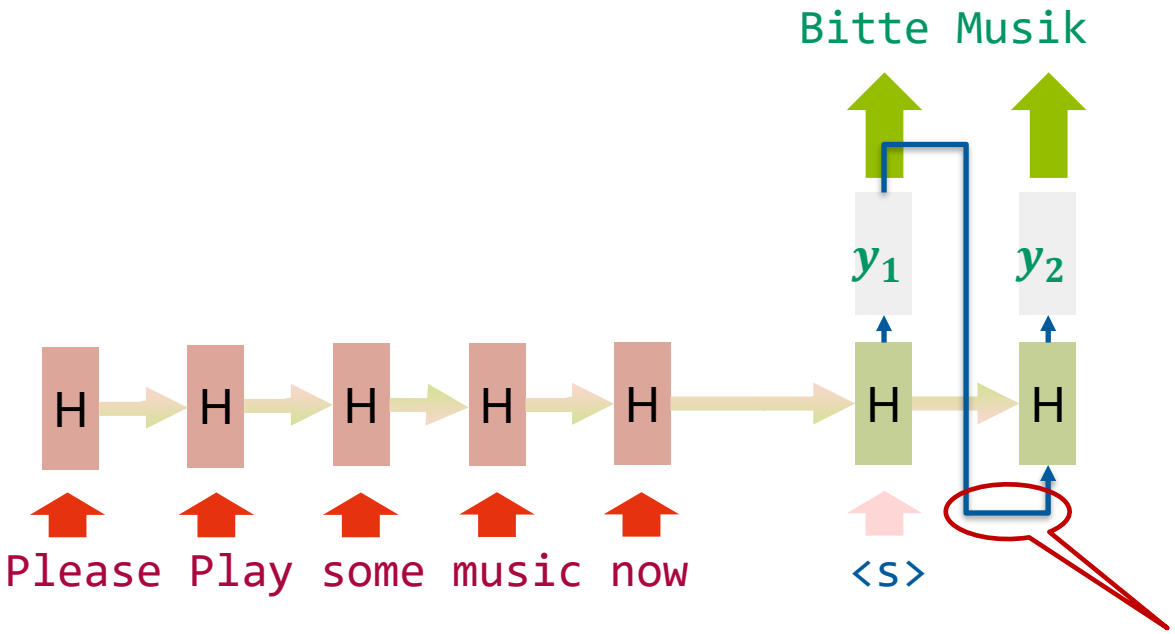


Example: NMT

Encoder-Decoder with Recurrent Neuronal Networks and Autoregression

Minimize $J = \sum_{t=1}^{T} -\log P(\boldsymbol{y_t}|\{y_i\}_{i<t}, x) = \quad J_1 \quad + \quad J_2$

J_1: negative log prob of „Bitte“



Input is previous prediction.

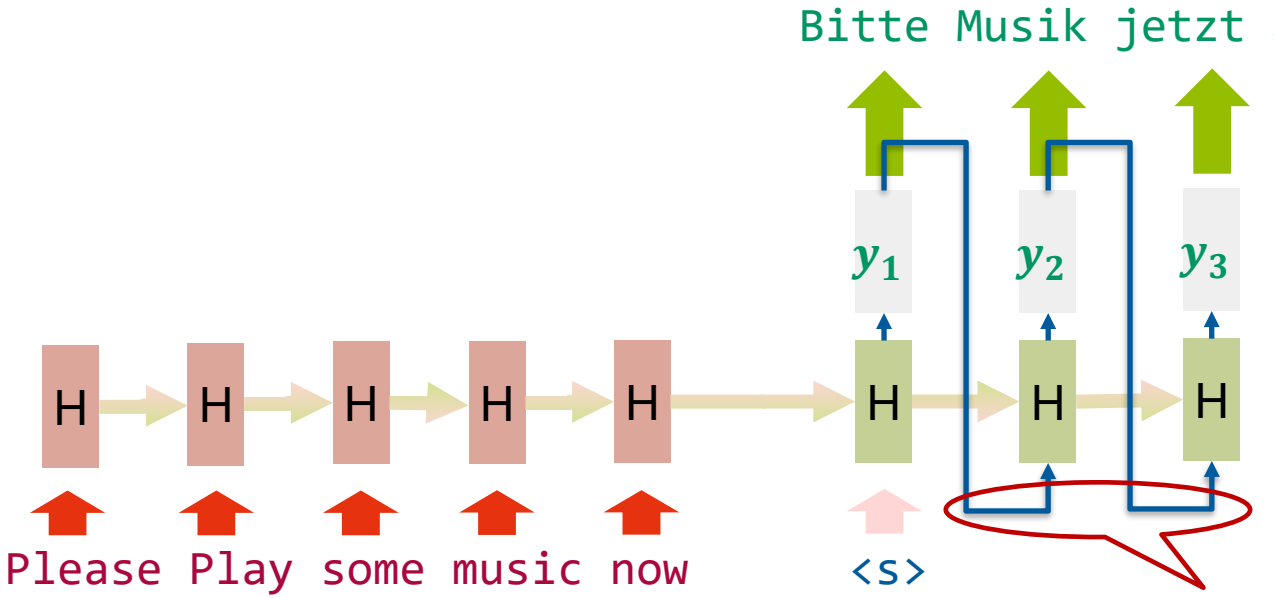
Example: NMT

Encoder-Decoder with Recurrent Neuronal Networks and Autoregression

Minimize $J = \sum_{t=1}^{T} -\log P(\boldsymbol{y_t}|\{y_i\}_{i<t}, x) =$ $J_1 + J_2 + J_3$

J_1: negative log prob of „Bitte“

J_3: negative log prob of „Jetzt“



This is called “autoregression”

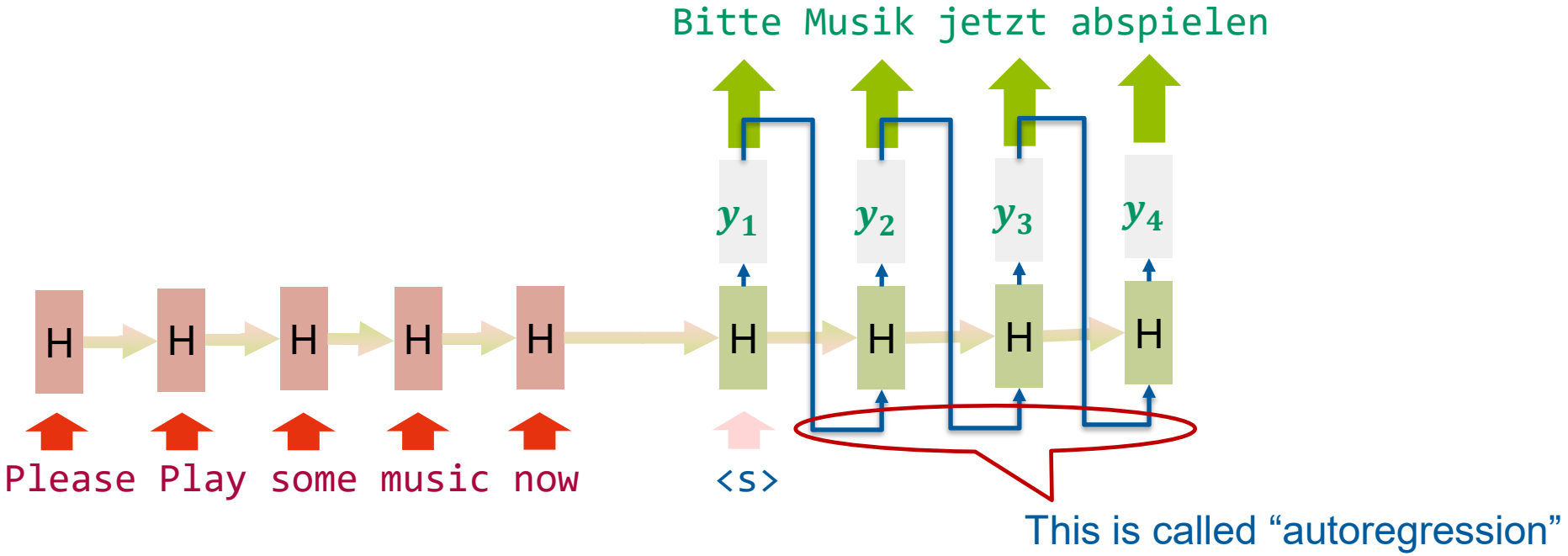
Example: NMT

Encoder-Decoder with Recurrent Neuronal Networks and Autoregression

Minimize $J = \sum_{t=1}^{T} -\log P(\boldsymbol{y_t}|\{y_i\}_{i<t}, \boldsymbol{x}) = \quad J_1 + J_2 + J_3 + J_4$

J_1: negative log prob of „Bitte“

J_3: negative log prob of „Jetzt“



Example: NMT

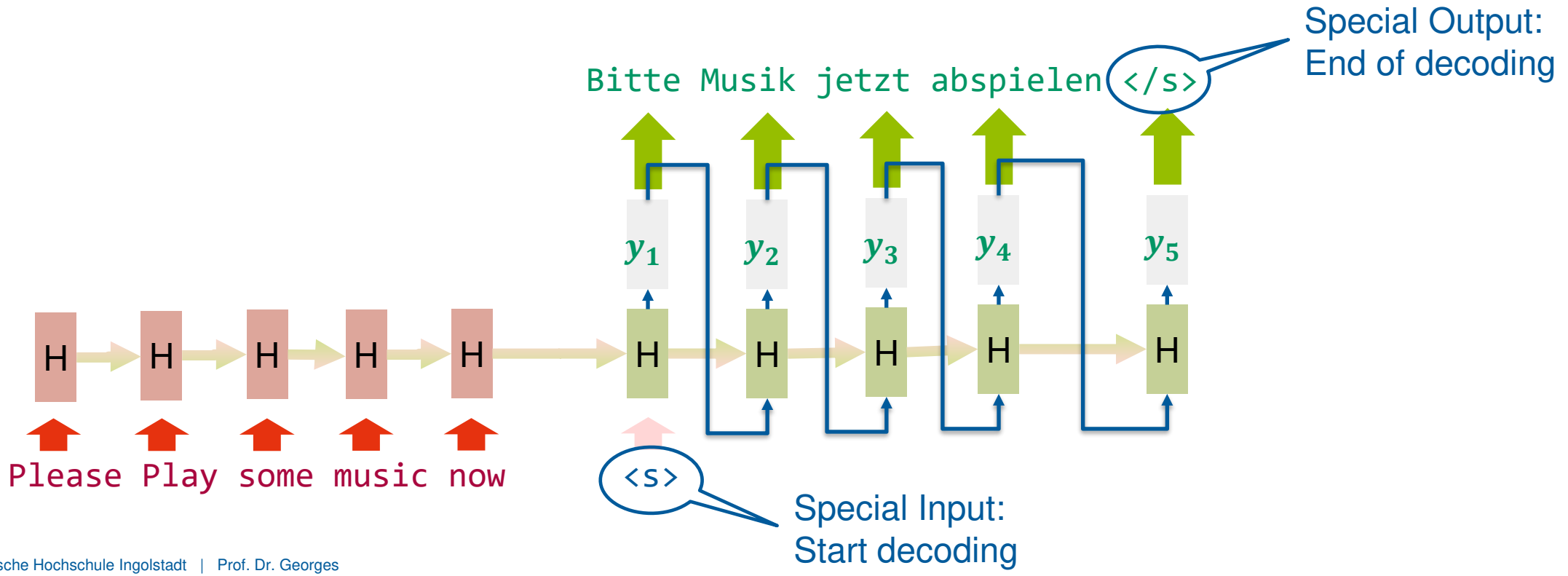
Encoder-Decoder with Recurrent Neuronal Networks and Autoregression

Minimize $J = \sum_{t=1}^{T} -\log P(y_t | \{y_i\}_{i<t}, x) = J_1 + J_2 + J_3 + J_4 + J_5$

J_1: negative log prob of „Bitte“

J_2: negative log prob of „Jetzt“

J_3: negative log prob of „</s>“



Example: NMT

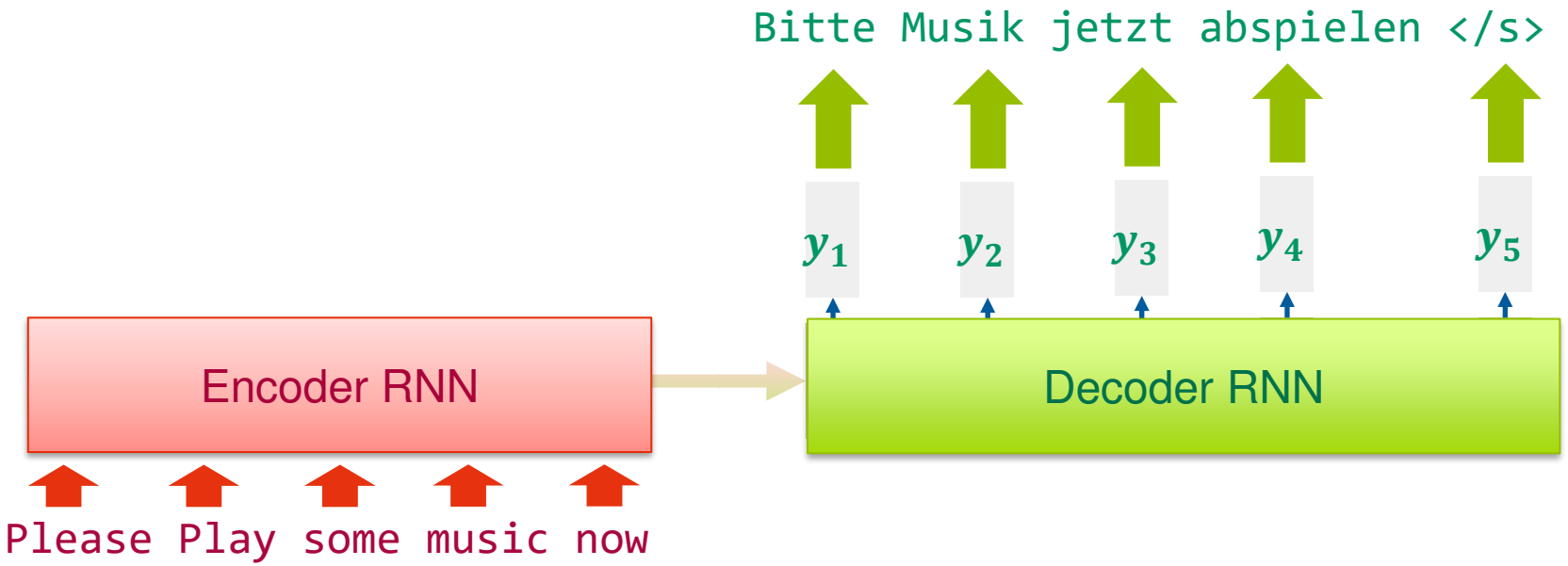
Encoder-Decoder with Recurrent Neuronal Networks and Autoregression

Minimize $J = \sum_{t=1}^{T} -\log P(\boldsymbol{y_t}|\{y_i\}_{i<t}, \boldsymbol{x}) = J_1 + J_2 + J_3 + J_4 + J_5$

J_1: negative log prob of „Bitte“

J_3: negative log prob of „Jetzt“

J_5: negative log prob of „</s>“



Example: NMT

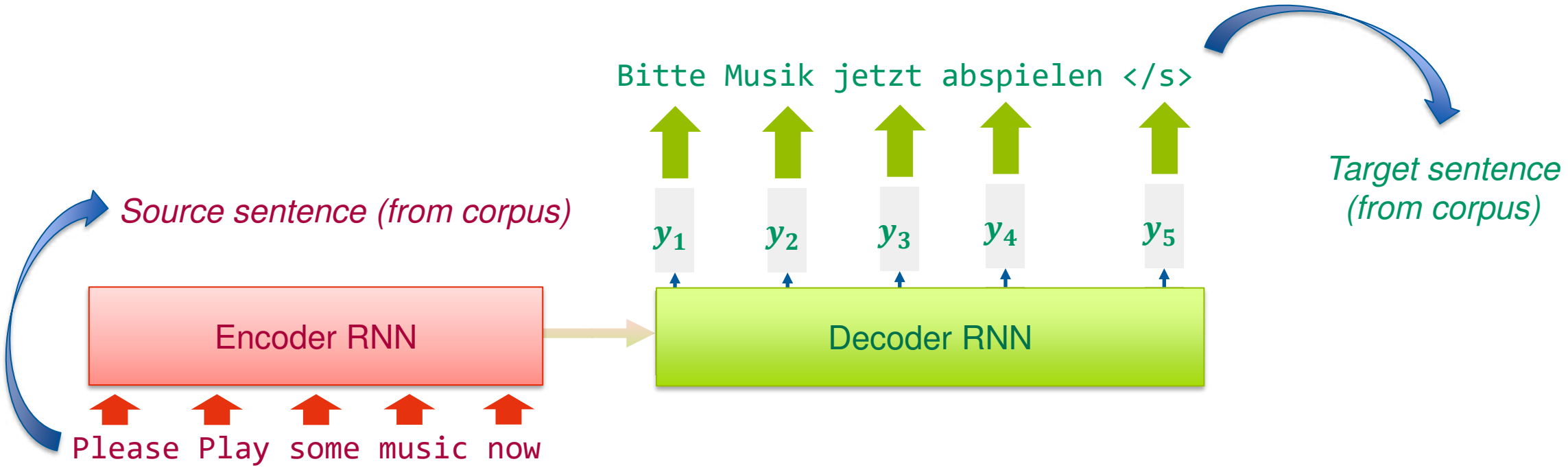
Encoder-Decoder with Recurrent Neuronal Networks and Autoregression

Minimize $J = \sum_{t=1}^{T} -\log P(y_t | \{y_i\}_{i<t}, x) = J_1 + J_2 + J_3 + J_4 + J_5$

J_1: negative log prob of „Bitte“

J_3: negative log prob of „Jetzt“

J_5: negative log prob of „</s>“



Example: NMT

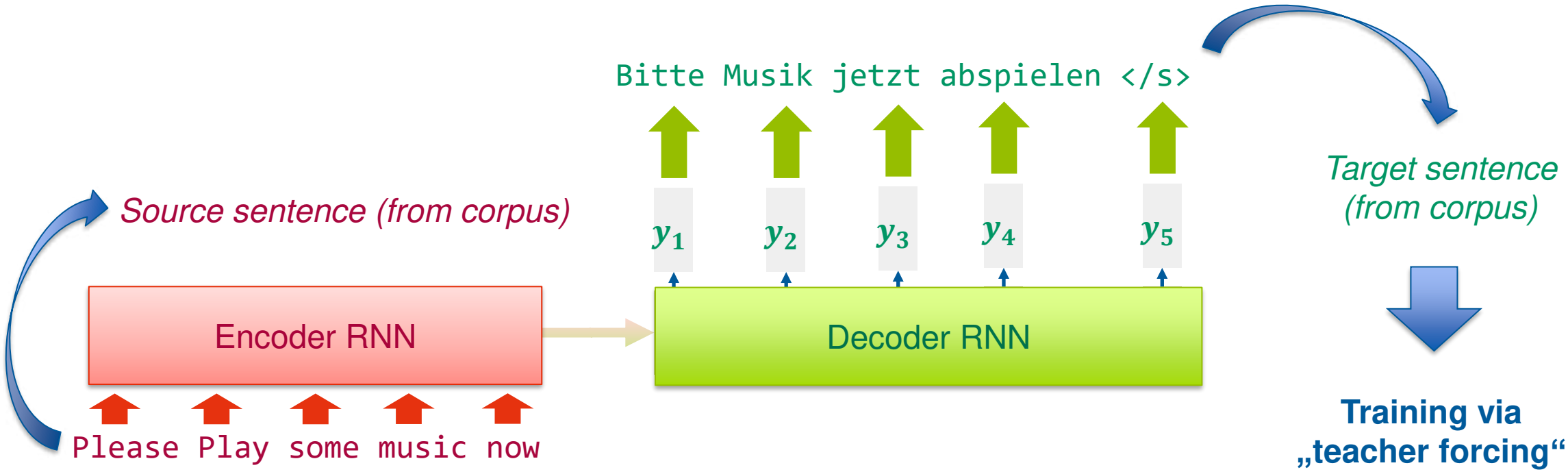
Encoder-Decoder with RNNs: “Teacher Forcing”

Minimize $J = \sum_{t=1}^{T} -\log P(\boldsymbol{y_t}|\{y_i\}_{i<t}, x) = J_1 + J_2 + J_3 + J_4 + J_5$

J_1: negative log prob of „Bitte“

J_3: negative log prob of „Jetzt“

J_5: negative log prob of „</s>“



“feeding observed sequence values (i.e. ground-truth samples) back into the RNN after each step”