RECURRENT NEURAL NETWORKS

Prof. Alexander Huth

4.14.2020

HOMEWORK 2

* Due today!

* If you haven’t turned it in yet, remember that the late policy is quite lenient (see syllabus)

PROJECT PROPOSALS

* Feedback will come to you sporadically over the next week

RECAP

* Linking artificial and biological neural networks

* In vision:
  * Train ANN to recognize objects
  * Use its representations to model visual cortex
  * This works pretty well (Eickenberg, Yamins)

RECAP

* How much does the *task* of the ANN matter?
  * Networks trained to classify images learn good representations for modeling brain
  * But random (untrained!) networks also work pretty well!

RECURRENT NEURAL NETWORKS

* Suppose there's a sequence of input output pairs $\{(x_t, y_t)\}_{t=1,2,\ldots,T}$

* & suppose there are sequential relationships, i.e. y_t is dependent not only on x_t but also on previous x

RECURRENT NEURAL NETWORKS

* Suppose there's a sequence of input output pairs $\{(x_t, y_t)\}_{t=1,2,\ldots,T}$

* To apply a simple neural network we could feed in all x and y in parallel

  * But this doesn't generalize to different sequence lengths, and would use a *huge* number of parameters

RECURRENT NEURAL NETWORKS

* **Recurrent neural networks (RNNs)** can solve this problem

* RNNs can solve these sequence problems *without* a huge number of weights

RECURRENT NEURAL NETWORKS

* (Drawing of basic RNN)

* (Drawing of RNN training using backprop through time, BPTT)

LANGUAGE

* Experimental setting:
  * A subject listens to language (e.g. a narrative story) while BOLD responses are recorded from cortex using fMRI

* We want to do **system identification**: build some model that predicts BOLD responses from the language stimuli

LANGUAGE

* To apply the “**system construction**” approach:

  * build an artificial neural network that solves **some task**

  * then use its representations (as a **linearizing transform**) to model the brain

* ***What task?***

LANGUAGE

* In vision: **object categorization** seems like a good task (modulo earlier discussion), because it requires the network to learn lots of things

* What is an equivalent for language?

  * Ideally something related to language **meaning**

LANGUAGE MODELS

* One solution may be **language models**, which have taken on a similar role in the NLP field to image classification models (e.g. AlexNet) in computer vision

* The task of a language model is to predict a word from its context

  * e.g. $P(w_i|w_1,...,w_{i-1})$

LANGUAGE MODELS

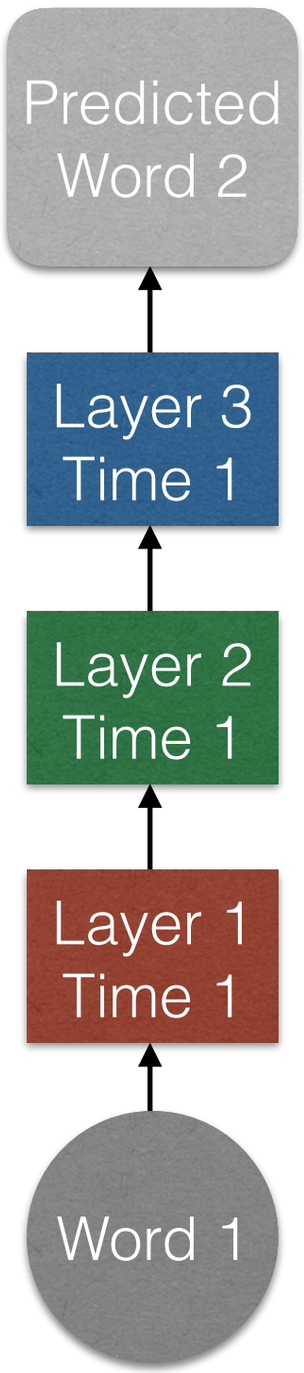
* Language models can use many different architectures

* One is a **recurrent neural network (RNN)**

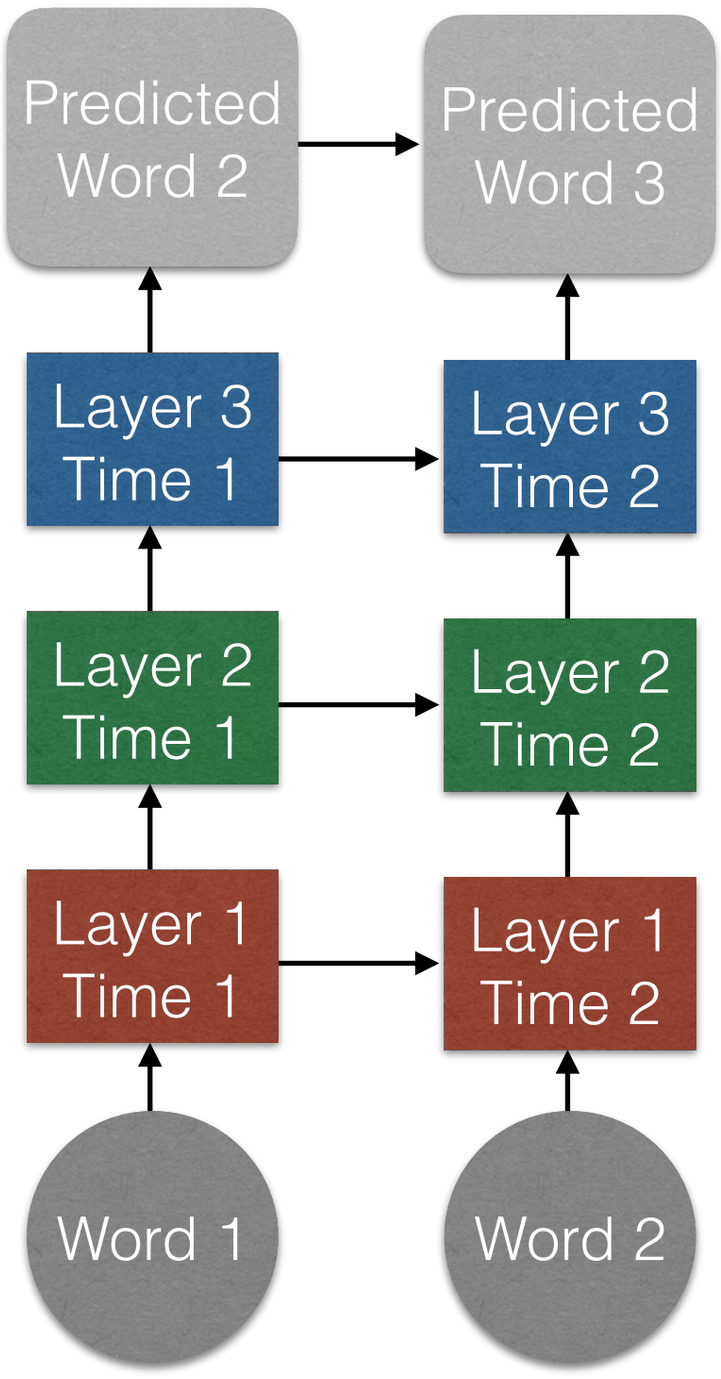
* In particular, a variant RNN called a **long short-term memory (LSTM) network**

  * (We'll talk talk about this network & how it works in detail on Tuesday)

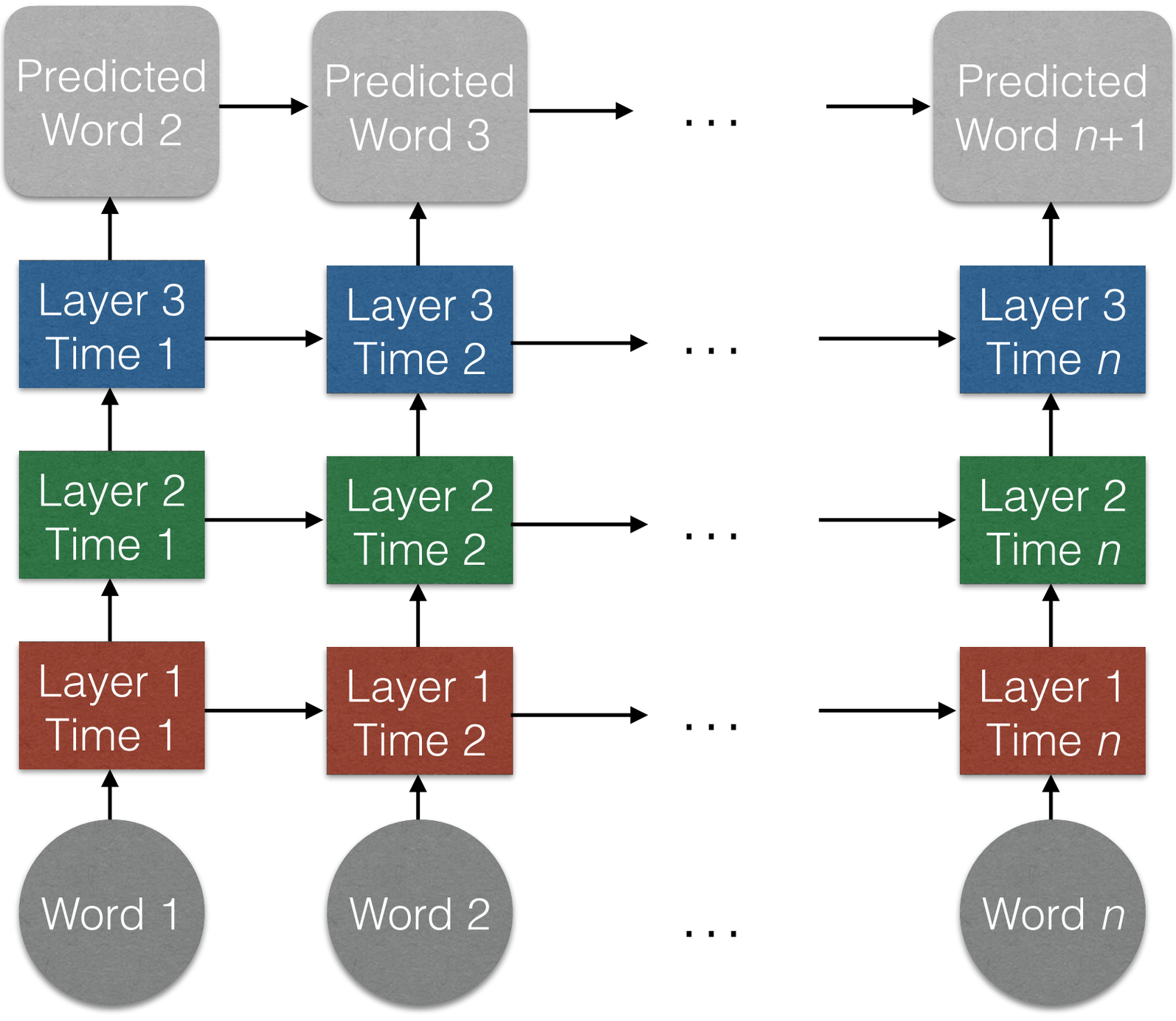
LSTM LANGUAGE MODEL



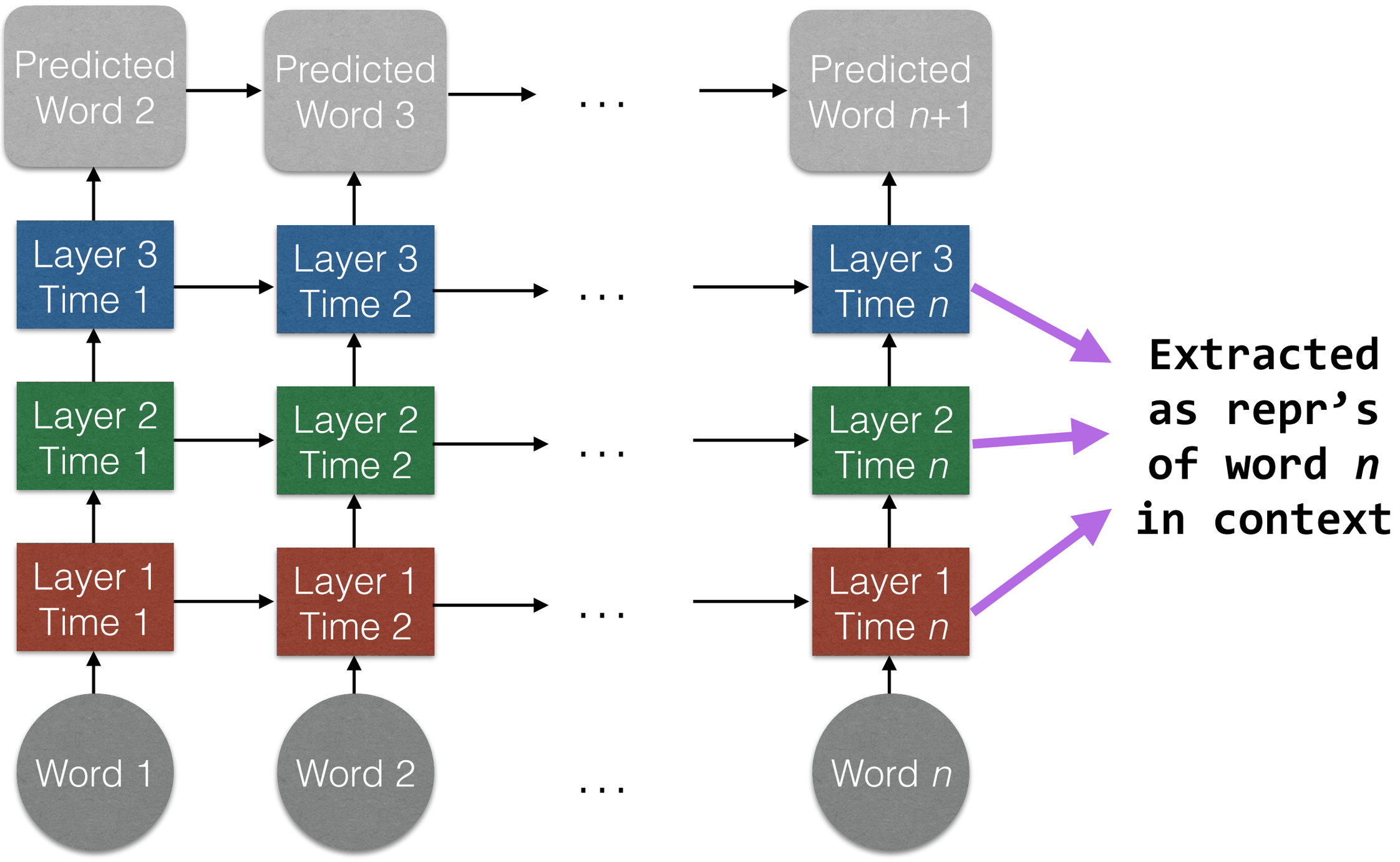
LSTM LANGUAGE MODEL



LSTM LANGUAGE MODEL



LSTM LANGUAGE MODEL



FMRI EXAMPLE USING LSTM LANGUAGE MODEL

* Jain & Huth (2018) Incorporating context into language encoding models for fMRI. *NeurIPS*

FMRI EXAMPLE USING LSTM LANGUAGE MODEL

* Approach:
  * Train LSTM language model on text
  * Do fMRI experiment
  * Use LSTM language model to extract features from fMRI stimuli
  * Build linearized system identification model using LSTM-derived features

FMRI EXAMPLE USING LSTM LANGUAGE MODEL

* In visual models different features can be extracted from different **layers**

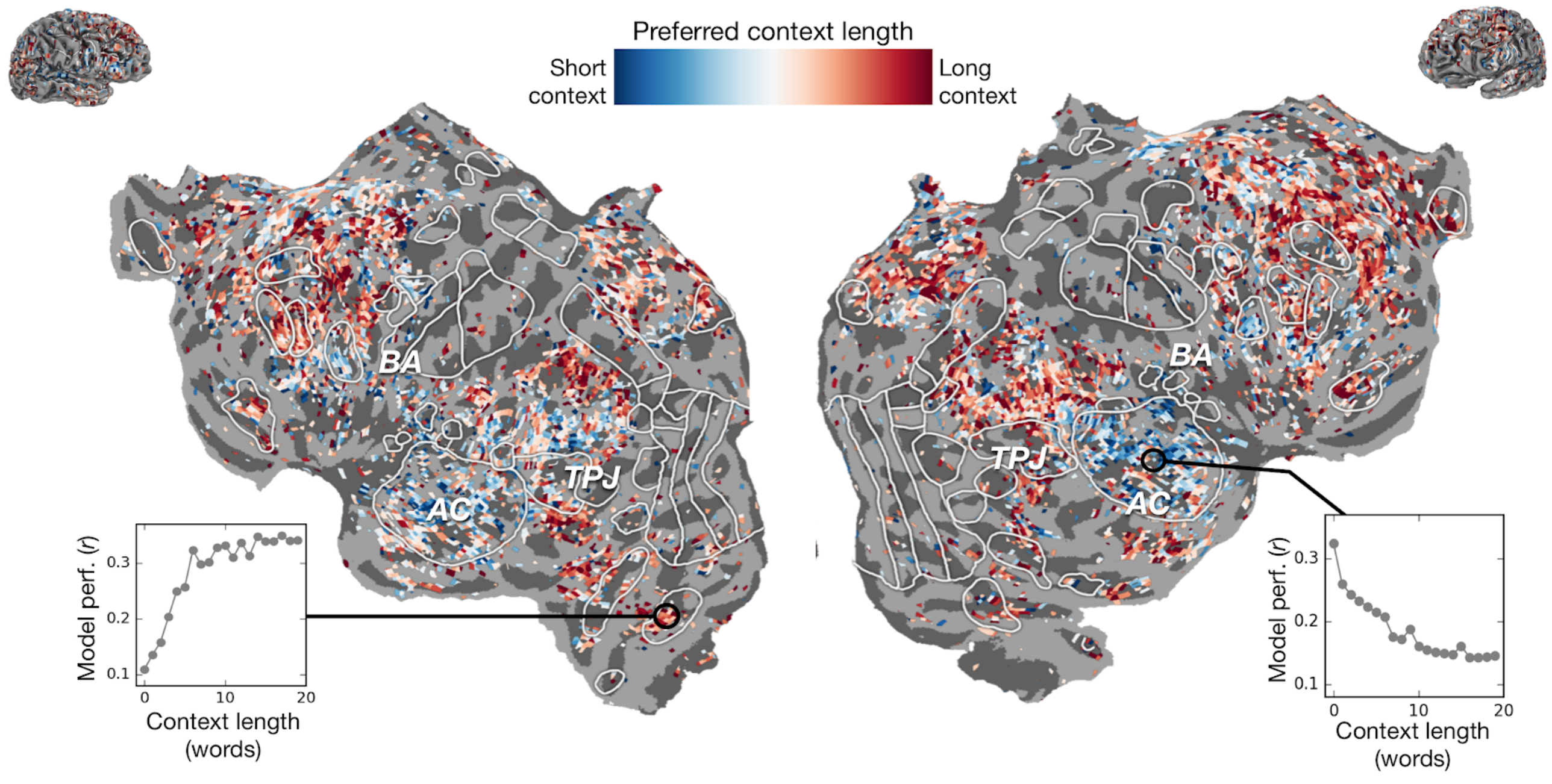
* In language models different features can be extracted from different **layers** *and* with different **amounts of context**

  * e.g. $P(w_i|w_{i-c},\ldots,w_{i-1})$

“context length”

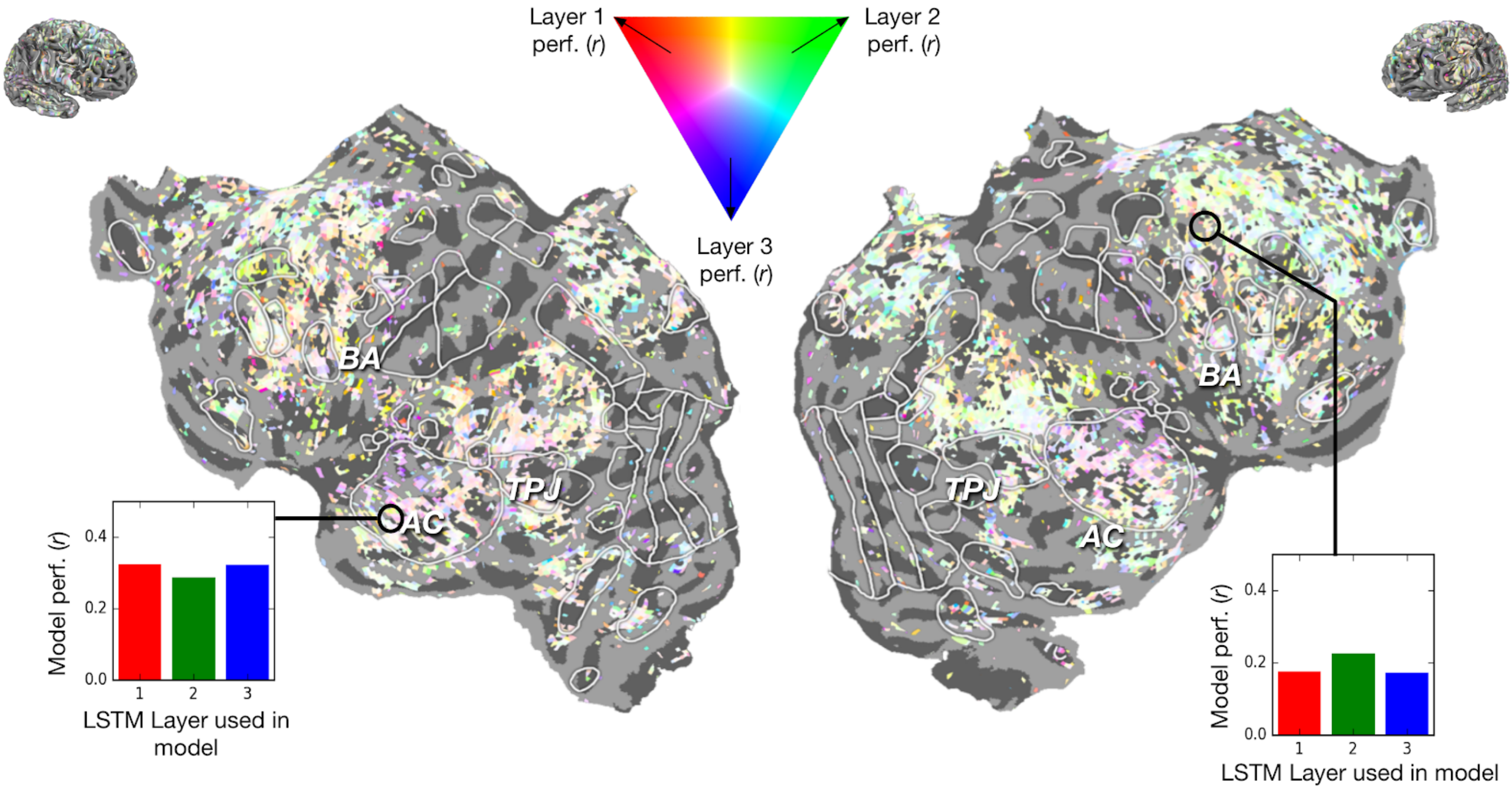
FMRI LSTM EXAMPLE

* Different brain areas prefer different amounts of context



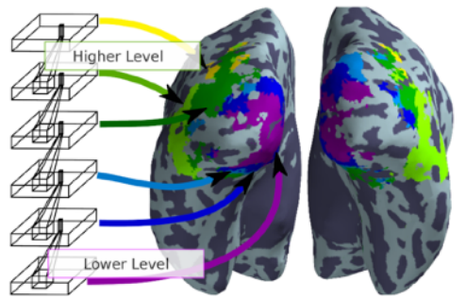
FMRI LSTM EXAMPLE

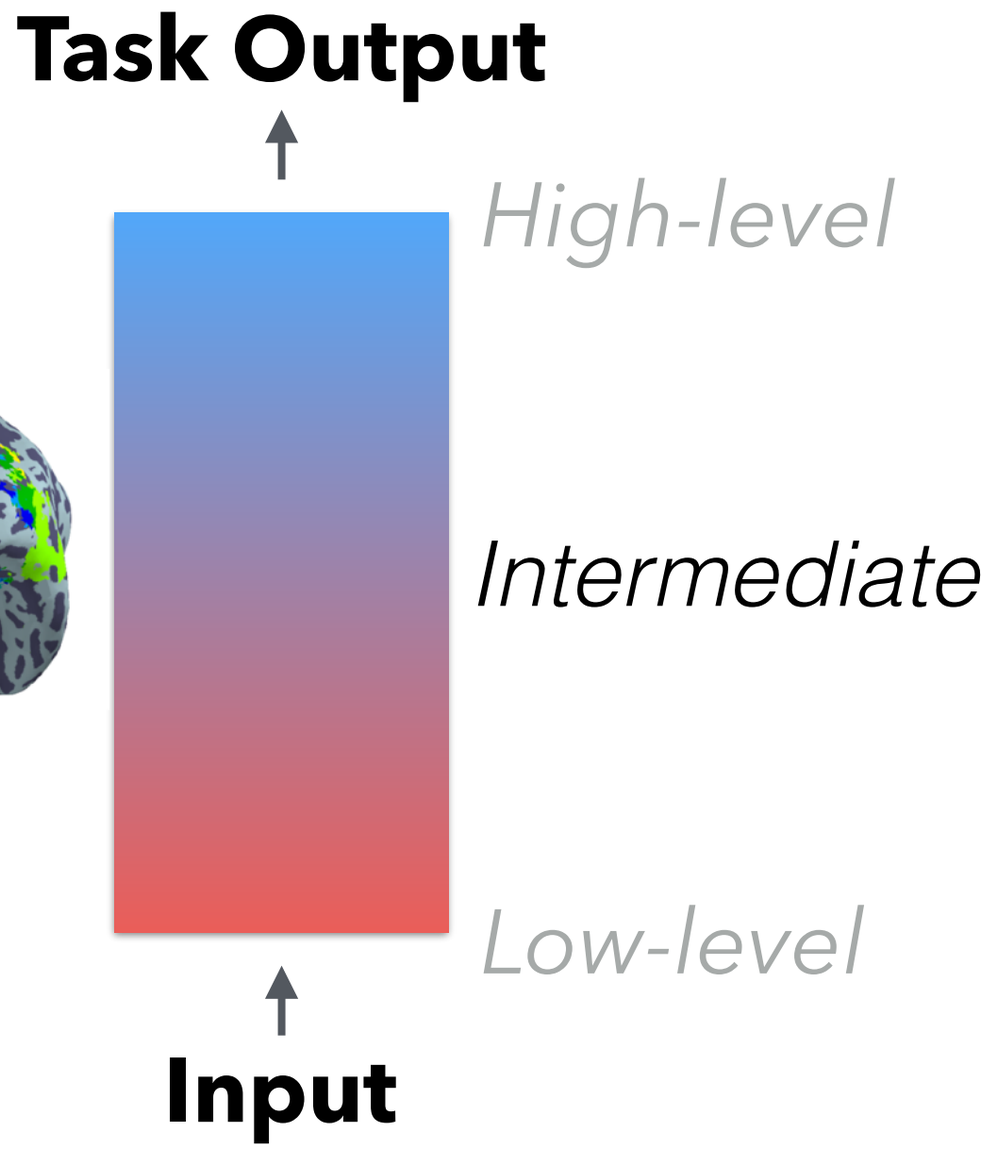
* But “layer preference” does not recapitulate known hierarchies, unlike Eickenberg, etc.



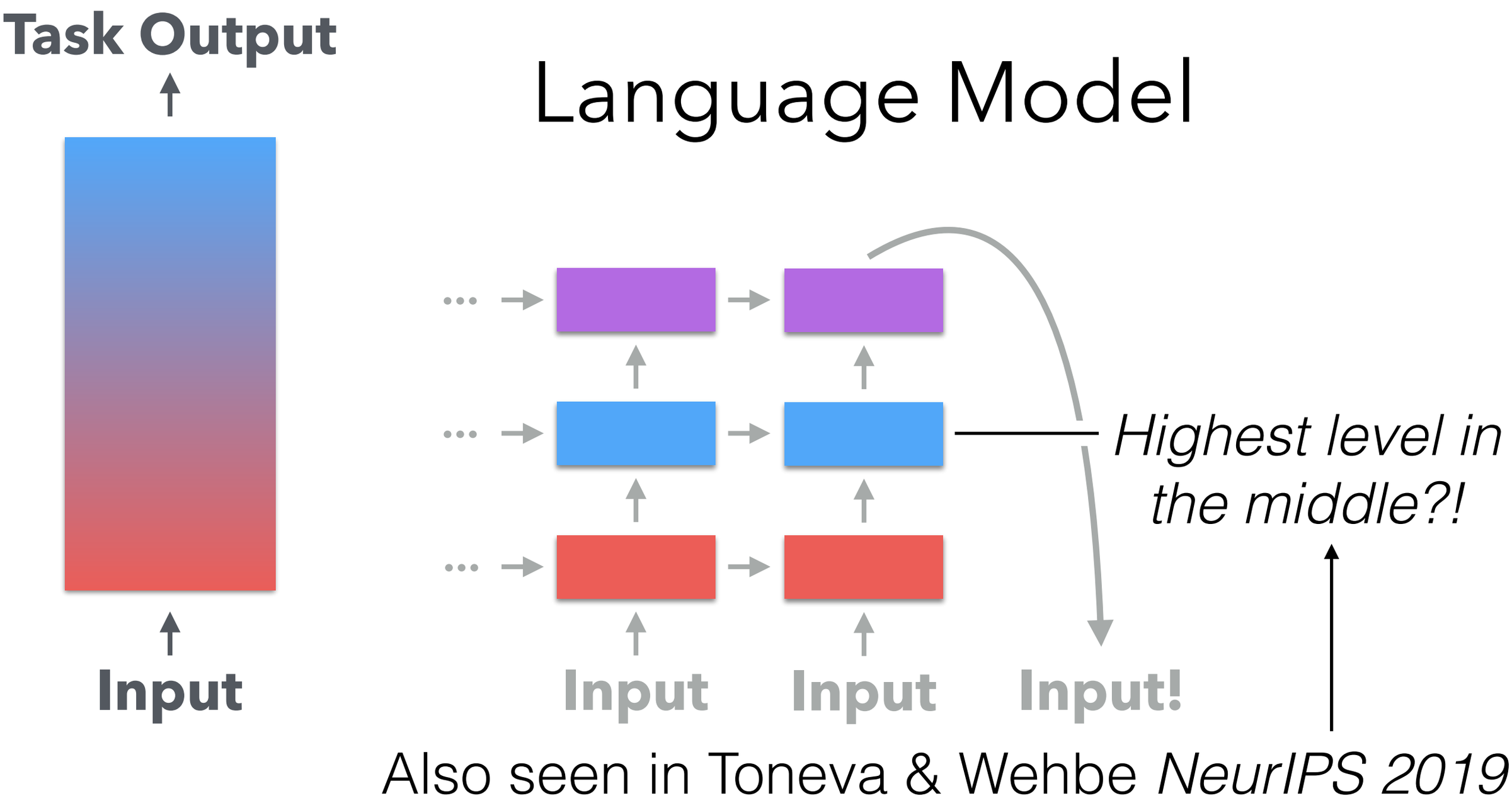
FMRI LSTM EXAMPLE

* In visual models, there is a clear “progression” of representations from low- to high-level





FMRI LSTM EXAMPLE



Also seen in Toneva & Wehbe *NeurIPS 2019*

NEXT TIME

* Other neural network architectures for language modeling: LSTMs & Transfomers