Experimental and Computational Methods in Linguistic Research

Spring 2025

Instructor: Sanghee Kim

Week 4

Agenda

- Prediction
- N400 effect
- Cloze probability
- Log probability (surprisal)
- Python `minicons` library
- R ggplot2 (for bar plots)

Install packages

```
# Comment out below if you already have the libraries installed
!pip install minicons
!pip install torch
!pip install matplotlib pandas
```

Then Runtime > Restart Session

Prediction and N400 effect

Making predictions

He spread the warm bread with ________.

Making predictions

He caught the pass and scored a touchdown. There was nothing he loved more than a good game of

____

Making predictions

He caught the pass and scored a touchdown. There was nothing he loved more than a good game of

football

Making predictions

He caught the pass and scored a touchdown. There was nothing he loved more than a good game of

monopoly

Prediction and the brain

- That “huh?” moment when encountering words that are semantically incompatible with the context or world knowledge
- .. can be captured through electrophysiological brain component!

EEG

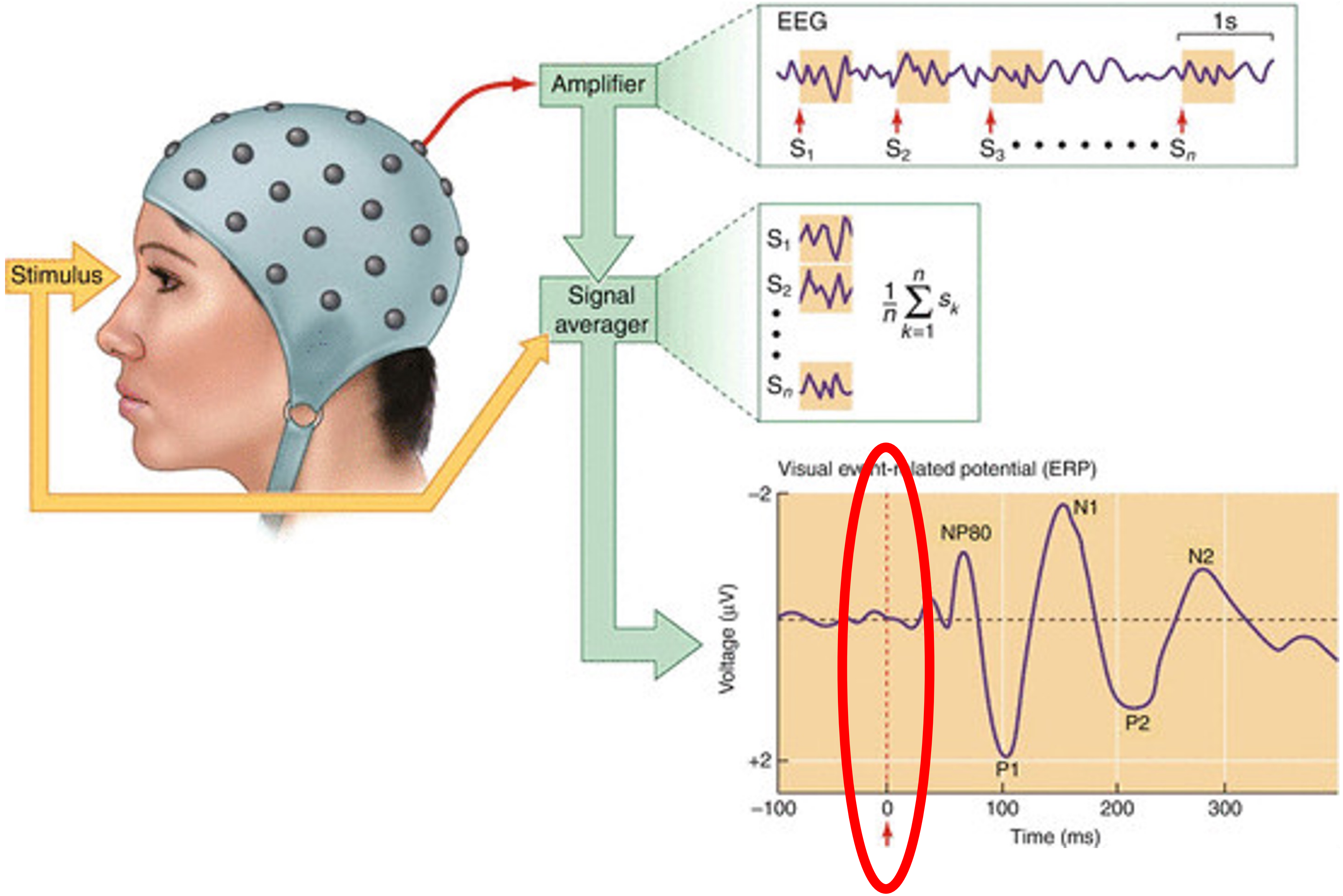
- Electroencephalography (EEG)
- Measures the electrical signal on the scalp

ERPs

- Since there are so many different sources of noise, it is impossible to detect the response from a single trial.

- Averaging across many trials removes the background EEG, leaving the event-related potential (**ERP**).

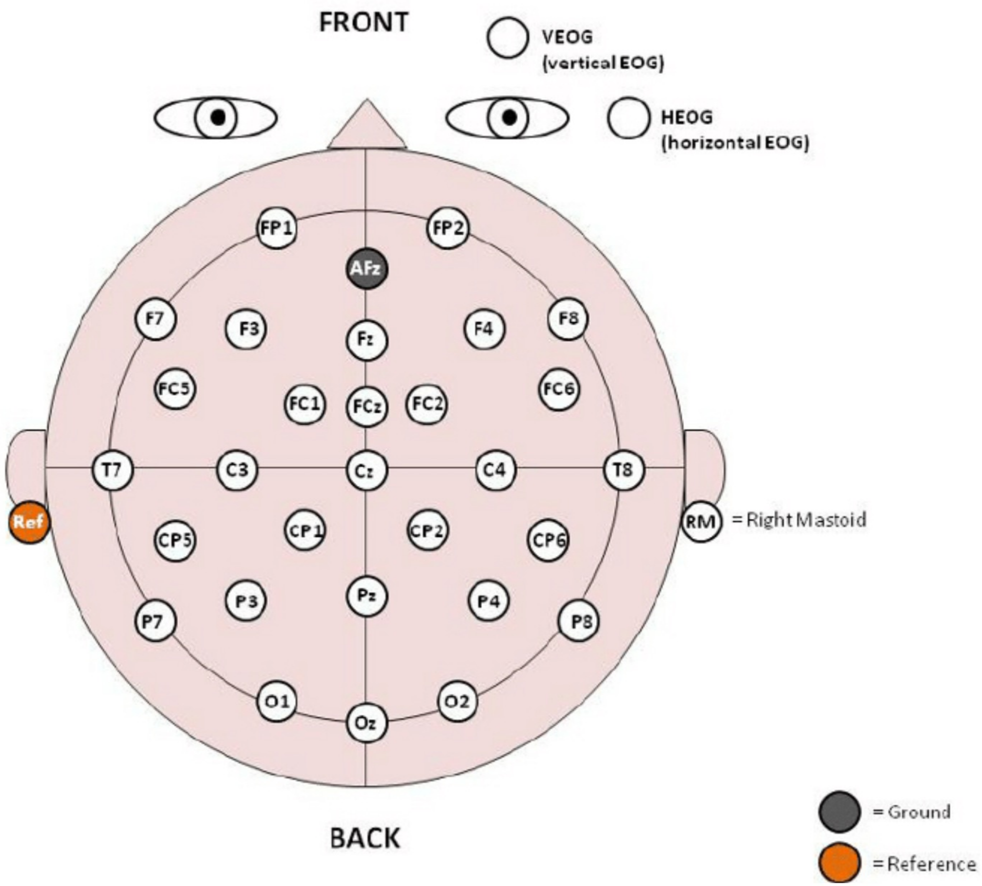
ERPs



Stimulus
Amplifier
EEG
1s
S1
S2
S3
Sn
Signal averager
$\frac{1}{n}\sum_{k=1}^{n} s_k$
Visual event-related potential (ERP)
−2
+2
NP80
N1
N2
P1
P2
Voltage (μV)
−100
0
100
200
300
Time (ms)

ERPs: 4 important parameters

- **Latency**: How soon after the stimulus is the effect found? (in milliseconds)
- **Polarity**: Pos/Neg-going waveform?
- **Amplitude**: How big is the effect? (in microvolts)
- **Topography:** Which area of the brain is the effect found?

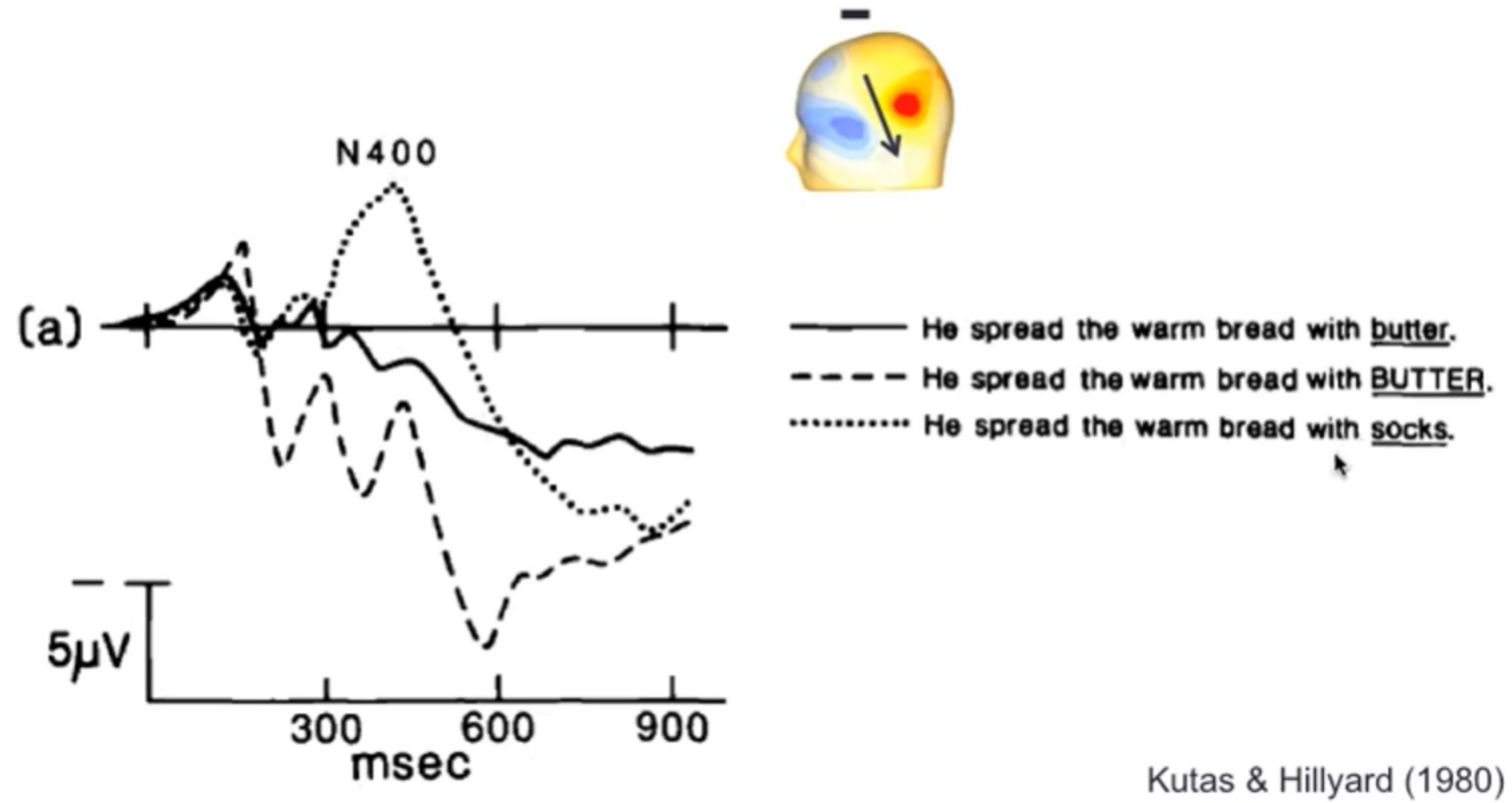


N400 effect

- Kutas and Hillyard (1980) discovered the N400 effect – signaling semantic anomaly.

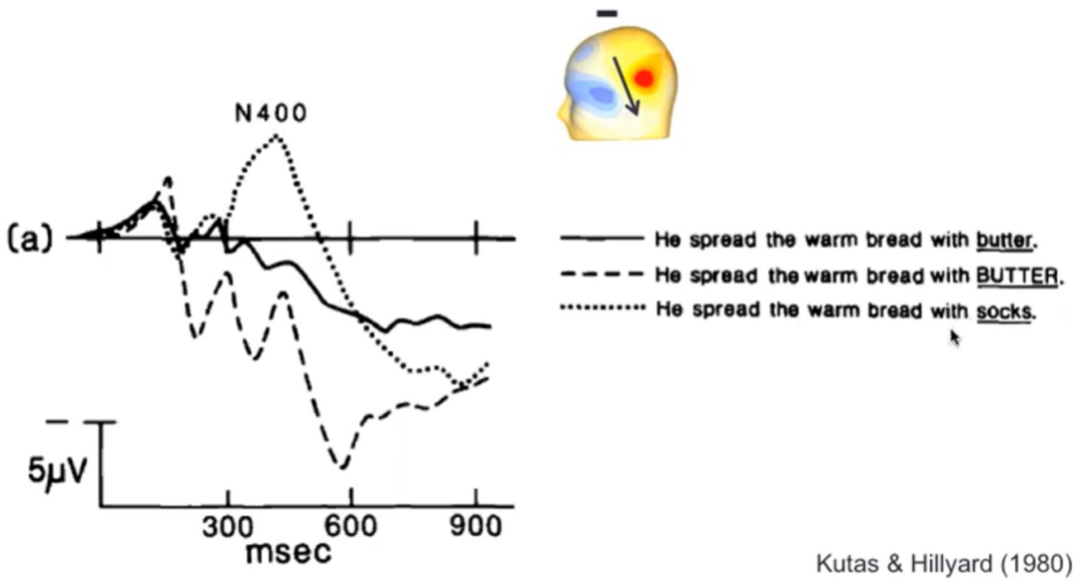
N400 effect

- Kutas and Hillyard (1980) discovered the N400 effect – signaling semantic anomaly.



Kutas & Hillyard (1980)

N400 effect

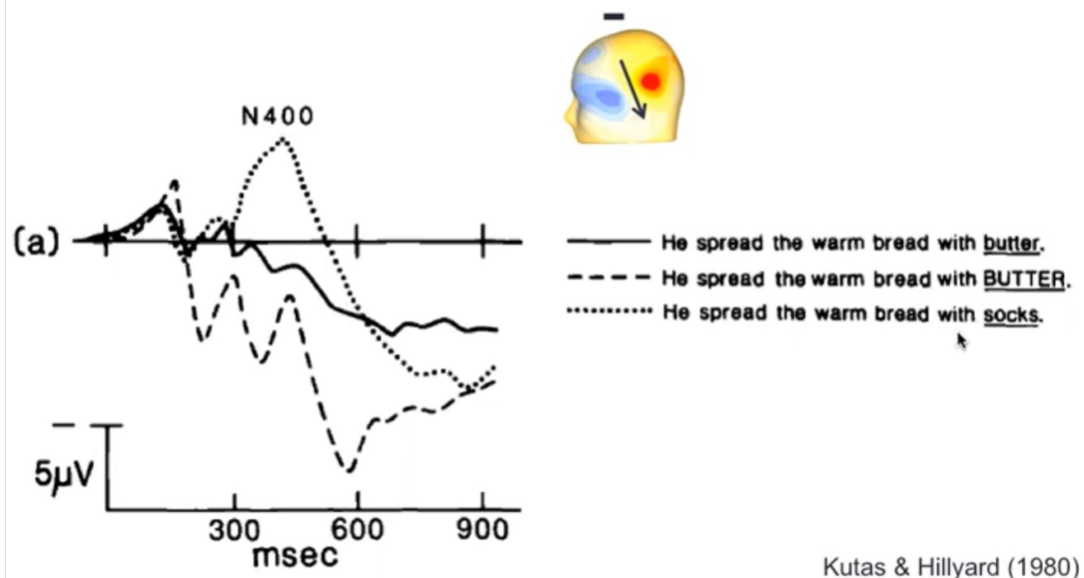


Kutas & Hillyard (1980)

Latency

- 250-500 (~400) ms after stimulus

N400 effect

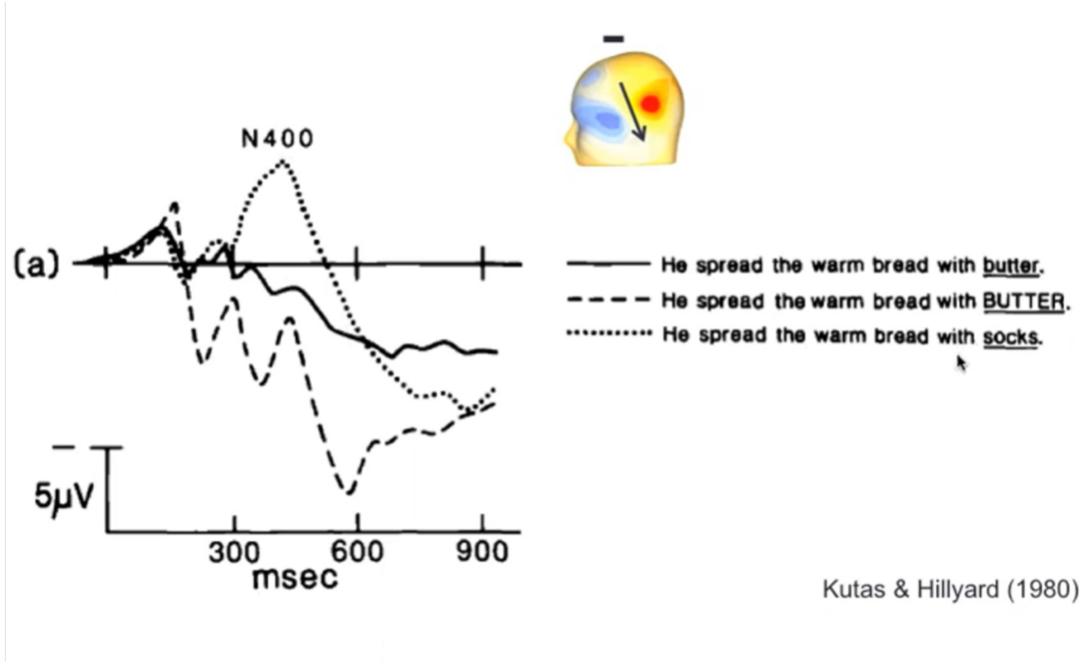


Kutas & Hillyard (1980)

Polarity

- Negative-going waveform
- (negative waveforms plotted upward, by convention)

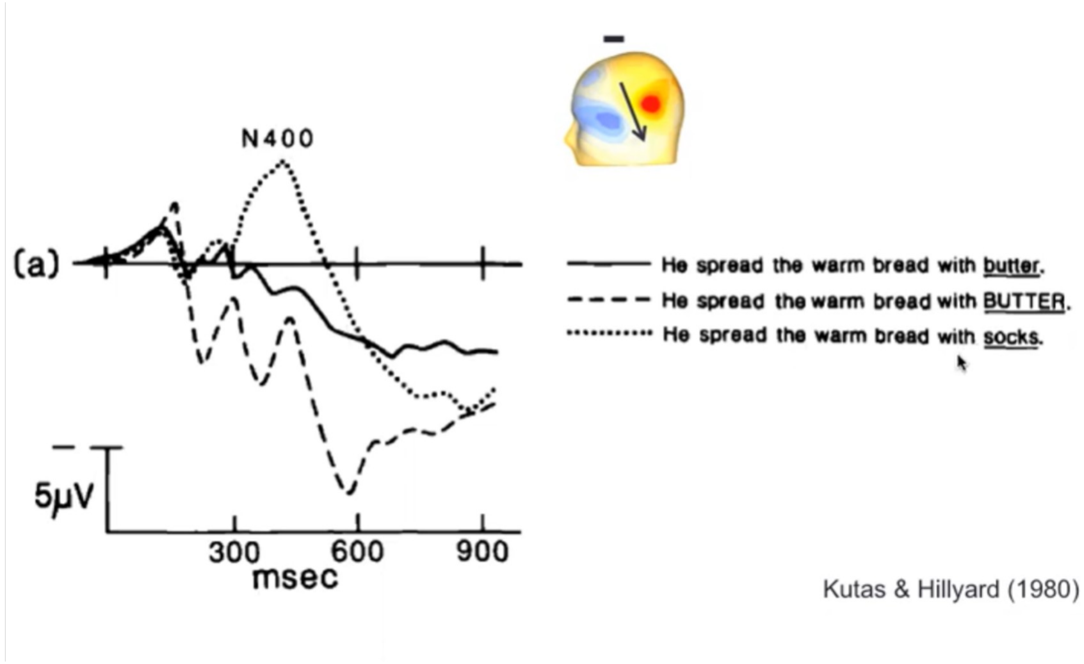
N400 effect



Amplitude

- Compare how much increased negative-going waveform relative to the expected word.
- Difference between the "butter" and "socks" condition.

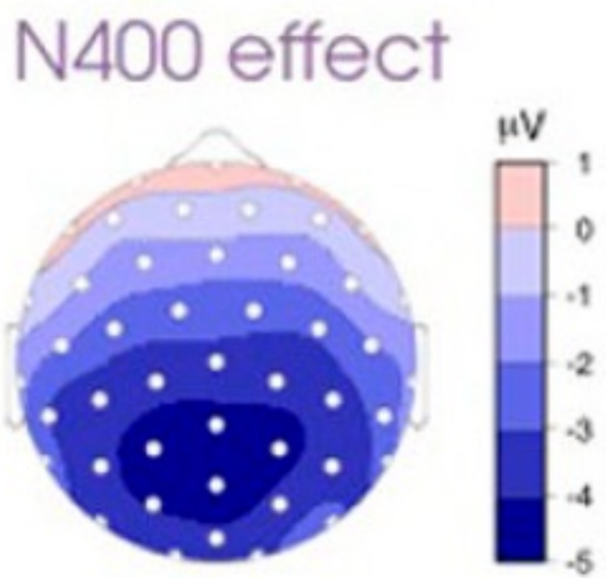
N400 effect



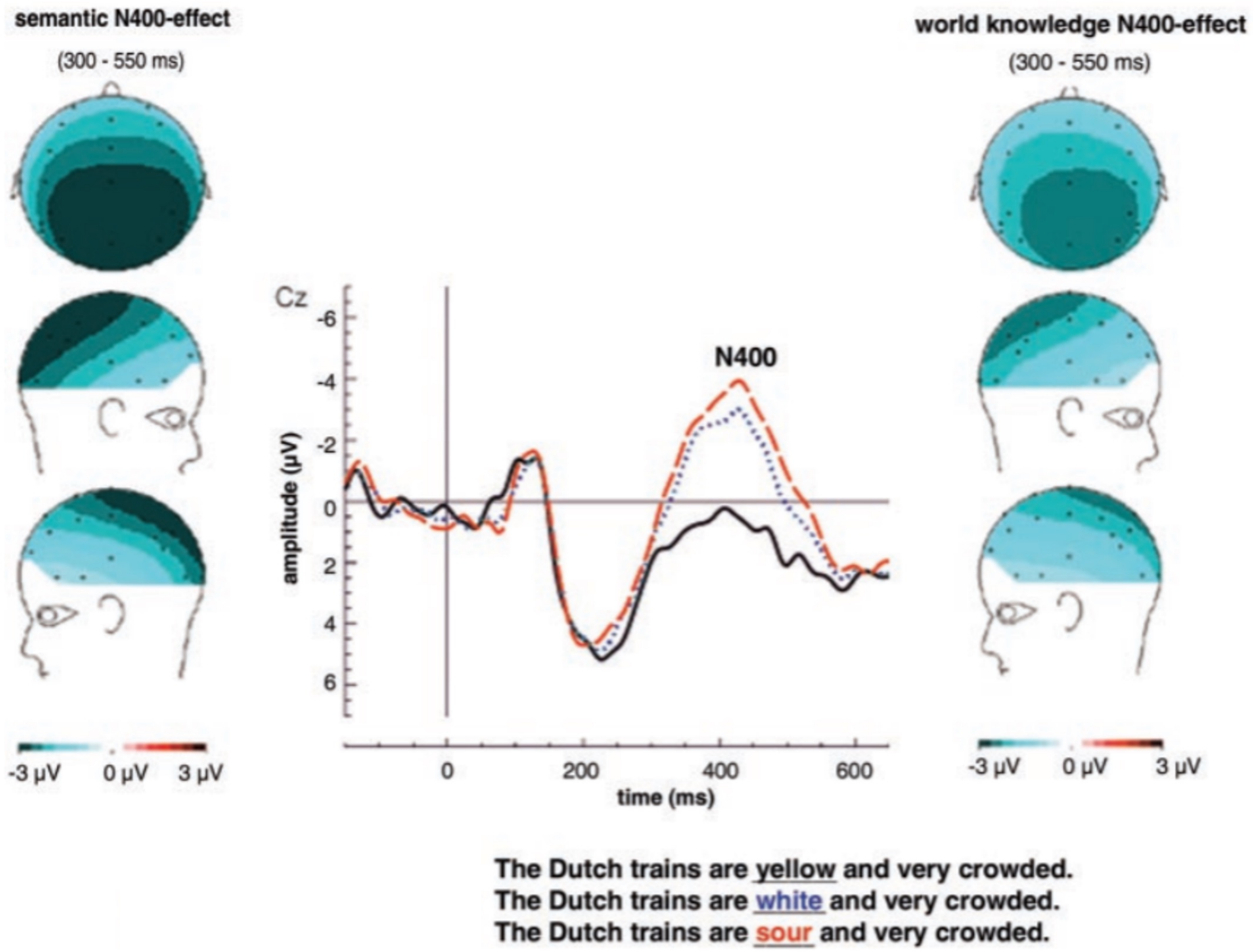
Kutas & Hillyard (1980)

Topography

- Centro-parietal sites, with a slightly right hemisphere bias

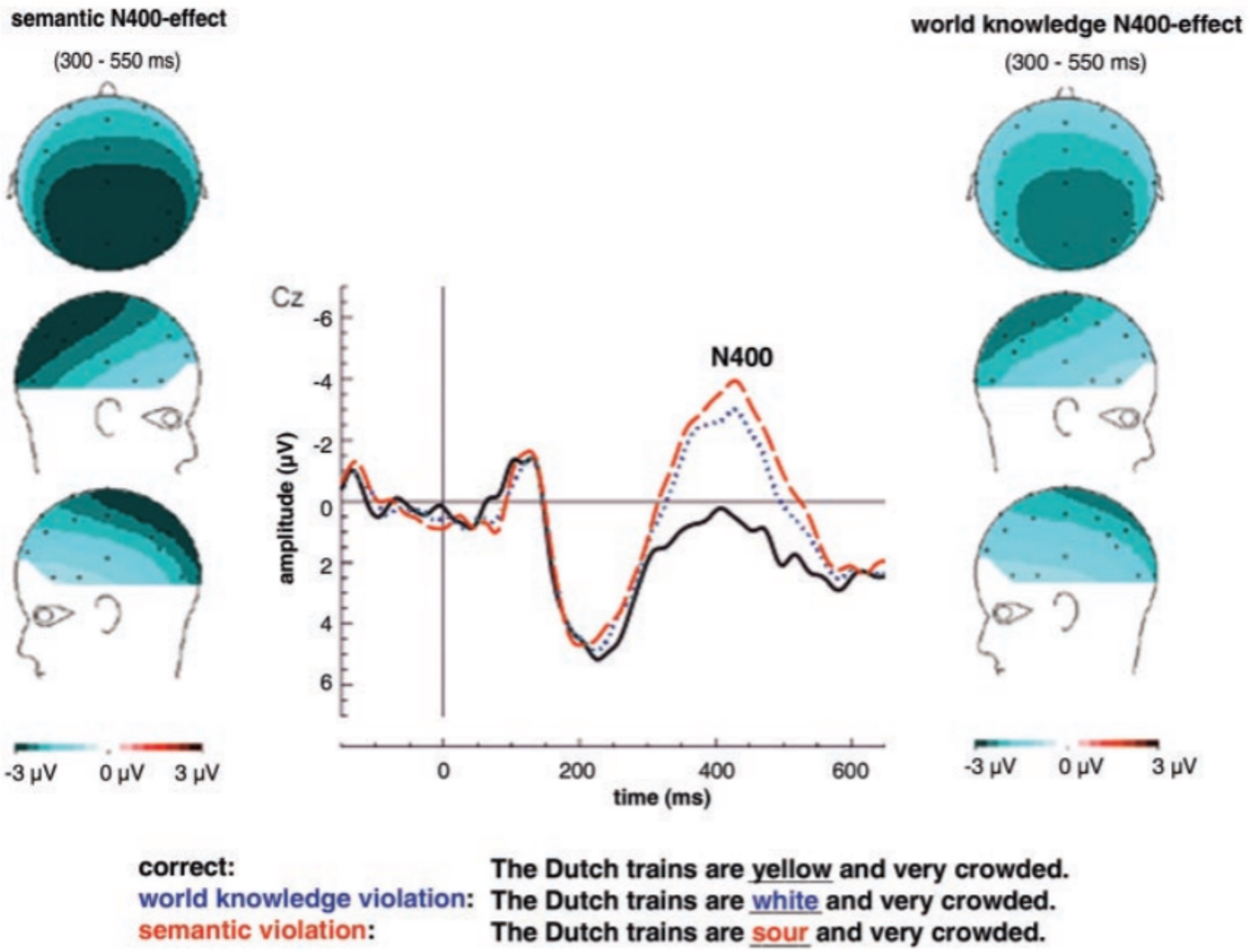


N400 effects



N400 effects

Interestingly, the N400 effect is also sensitive to world knowledge (Hagoort et al., 2004).



Summary

- Lexical access happens around 250 – 500 ms, and this is indexed by the N400 effect.
- The N400 effect occurs whenever some predictions/expectations (semantic and world knowledge) are violated.

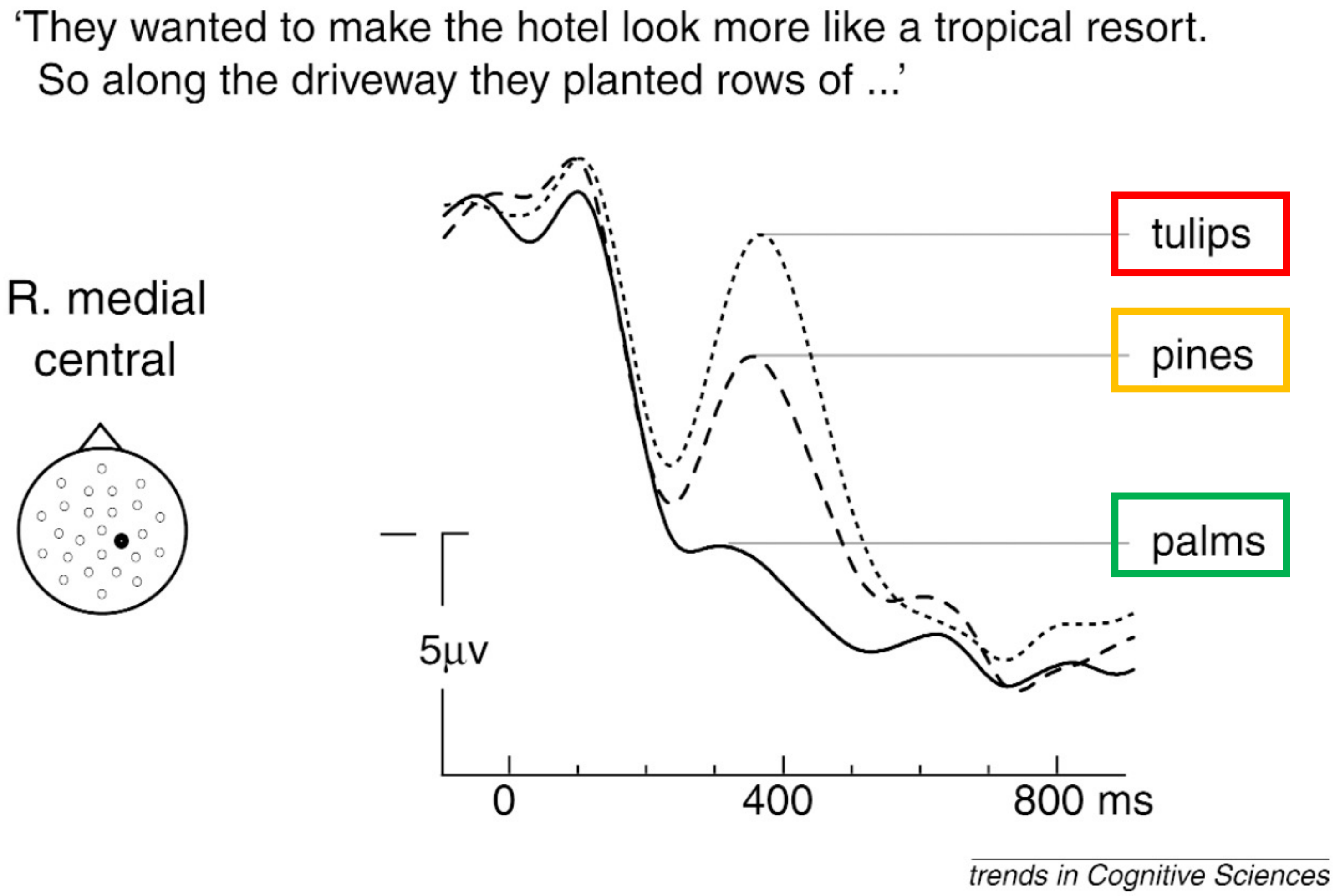
Cloze probability

N400 effect

‘They wanted to make the hotel look more like a tropical resort.
So along the driveway they planted rows of ...’

Kutas & Federmeier (2000) TiCS. Fig. 3.

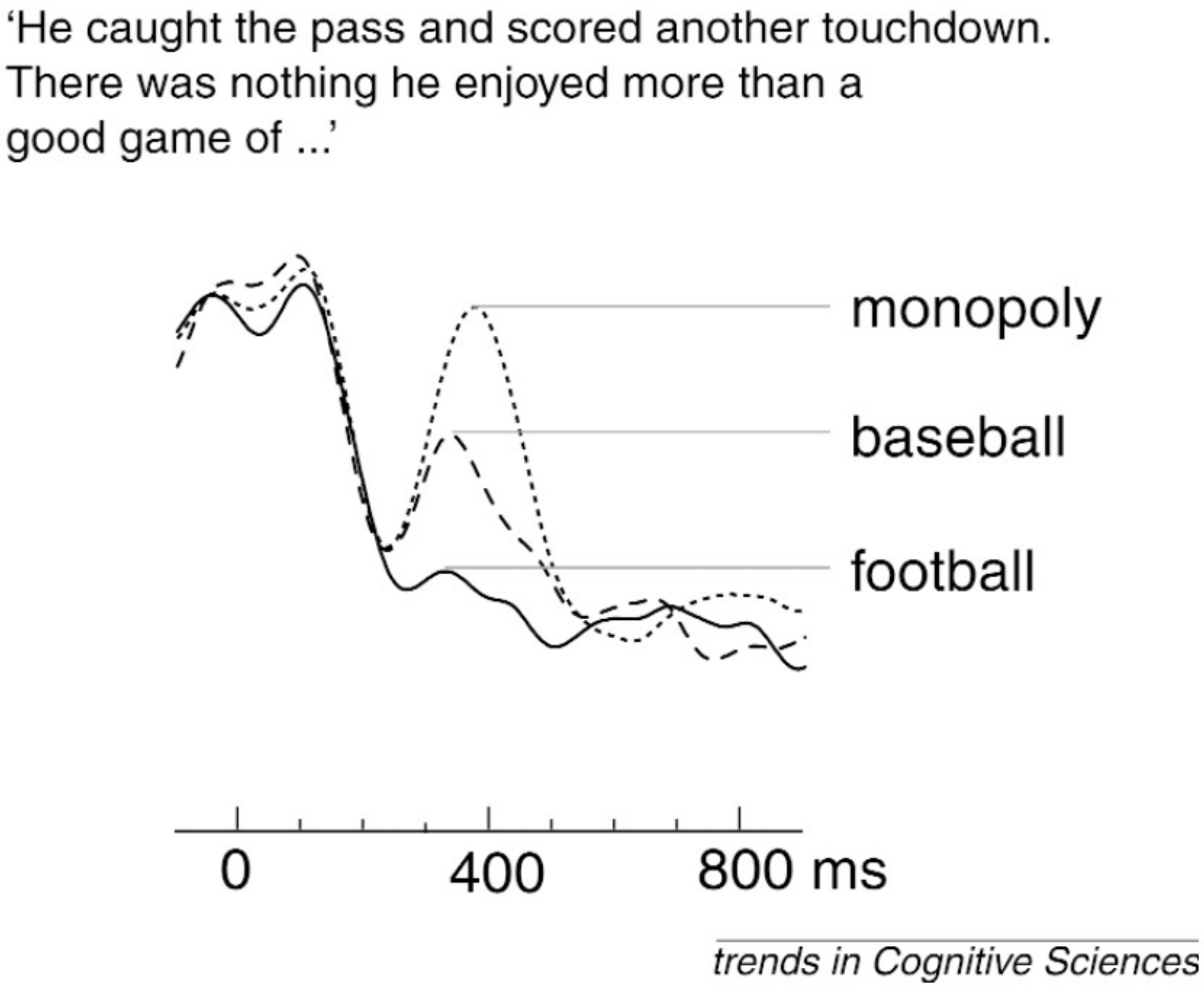
N400 effect



How do we create these three groups of words?

Kutas & Federmeier (2000) TiCS. Fig. 3.

N400 effect



How do we create these three groups of words?

Kutas & Federmeier (2000) TiCS. Fig. 3.

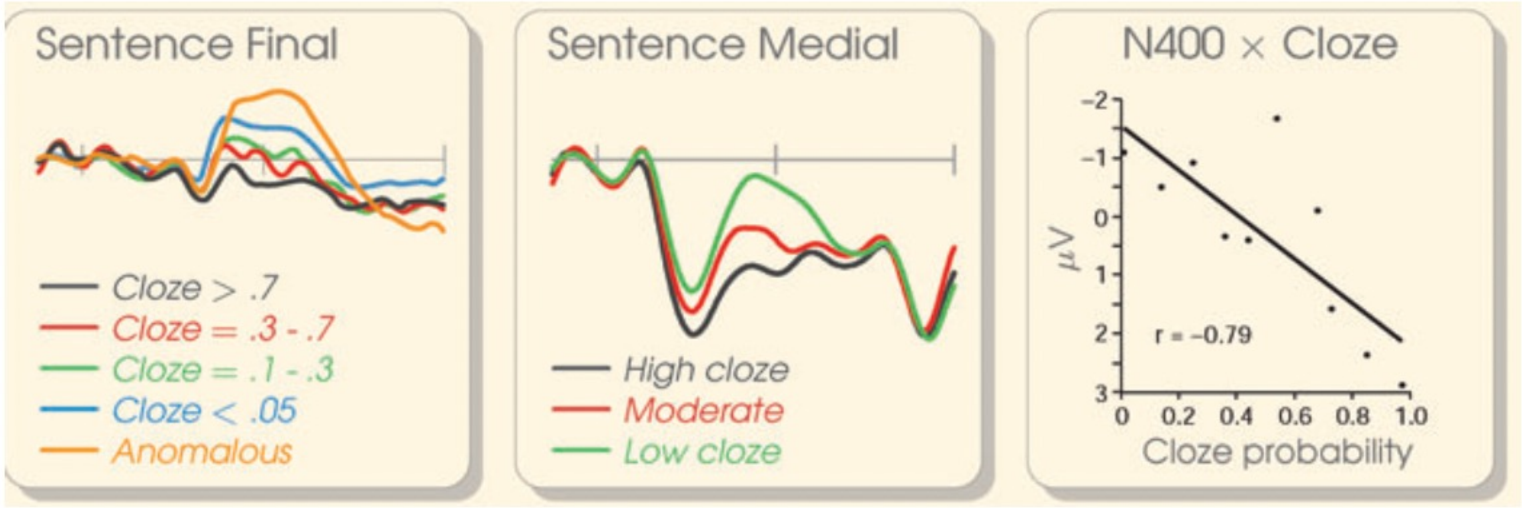
Cloze task

- **Cloze**: the percentage of respondents supplying a particular continuation for a context in an offline norming task (Taylor, 1953).
  - "I had coffee with cream and __."
  - E.g., 92/100 people responded: *sugar* (cloze probability = 0.92).

- (see the assigned reading, Kutas et al., 2011, for more.)

Cloze probability and N400 effect

- N400 amplitudes and cloze probabilities inversely correlated.
  - The higher a word's cloze probability, the smaller its N400 amplitude.

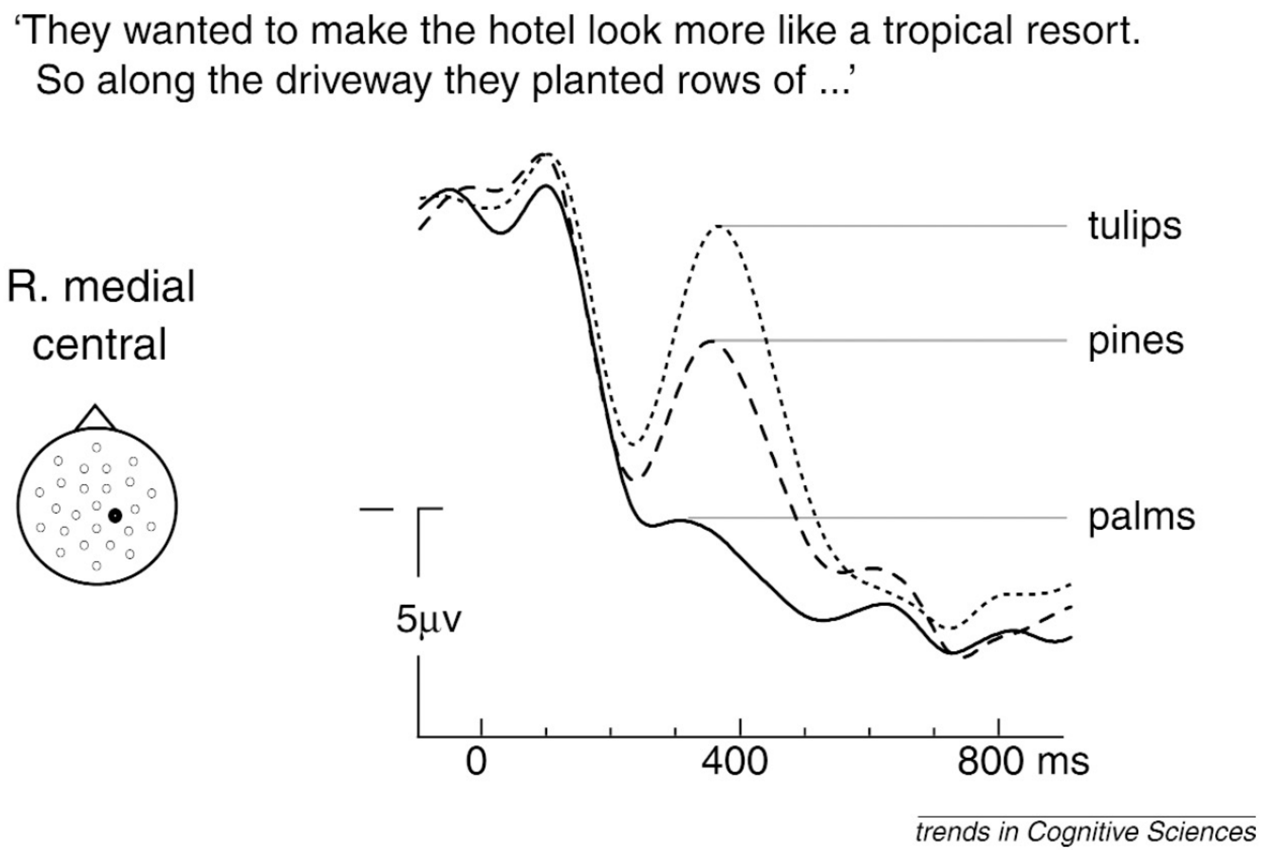


Kutas & Federmeier (2010)

Cloze probability

- We call it cloze “probability,” but is it just about the likelihood of the word (or string) appearing given prior context?
- .. or something more than this?
- How do we test this hypothesis?

Federmeier & Kutas (1999)



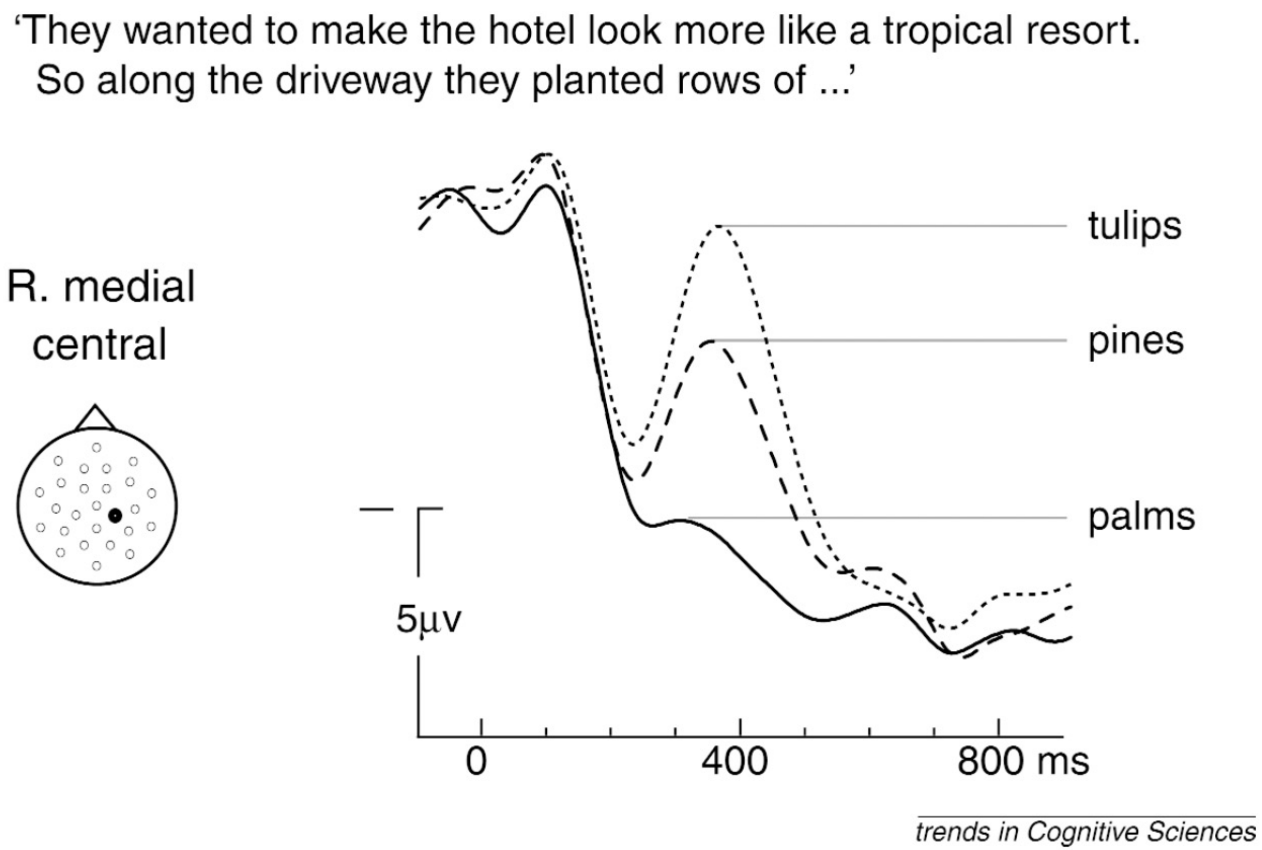
Cloze probability

- palms / pines / tulips
  0.74 / < 0.05 / < 0.05

Category membership

- palms / pines / tulips
  [tree] / [tree] / [flower]

Federmeier & Kutas (1999)



Cloze probability

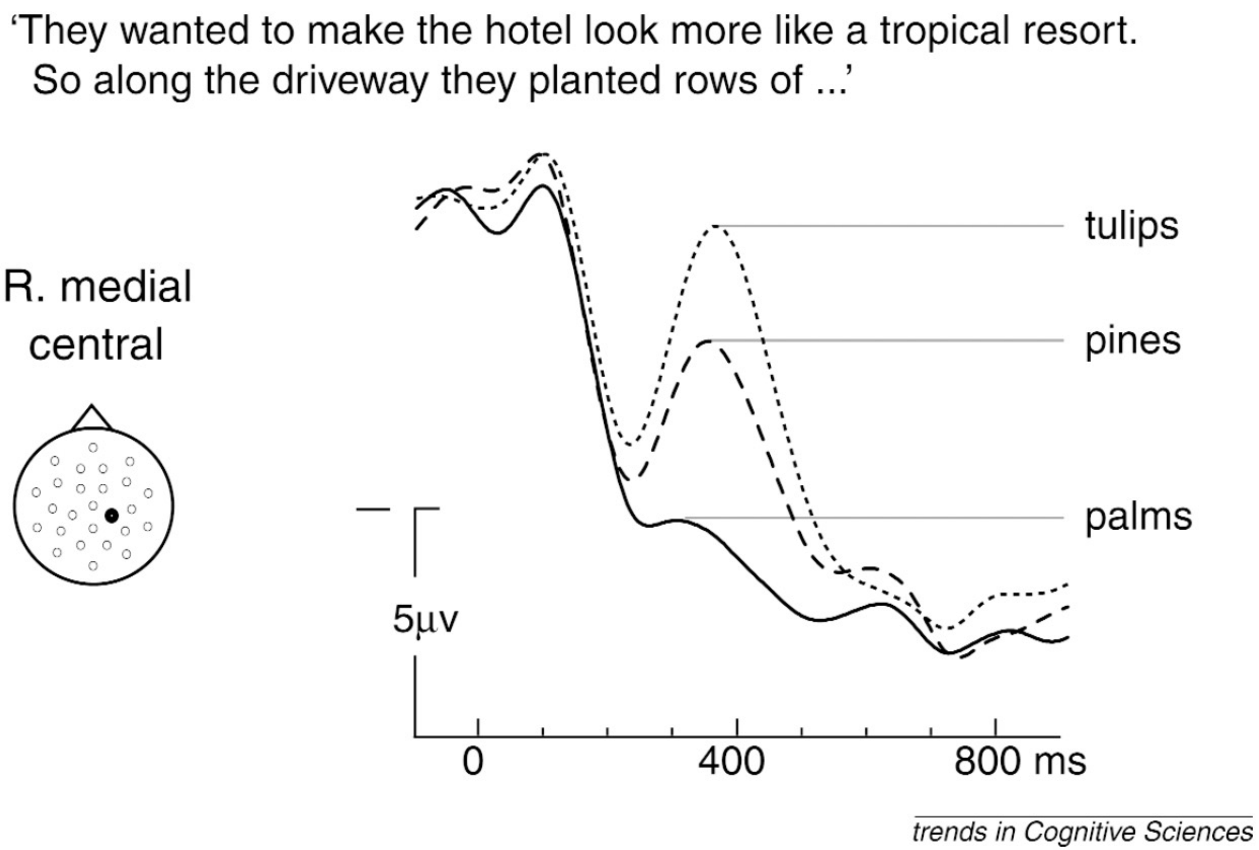
- palms / pines / tulips
  0.74 / < 0.05 / < 0.05

Category membership

- palms / pines / tulips
  [tree] / [tree] / [flower]

Unexpected; within category

Federmeier & Kutas (1999)



Cloze probability

- palms / pines / tulips
  0.74 / < 0.05 / < 0.05

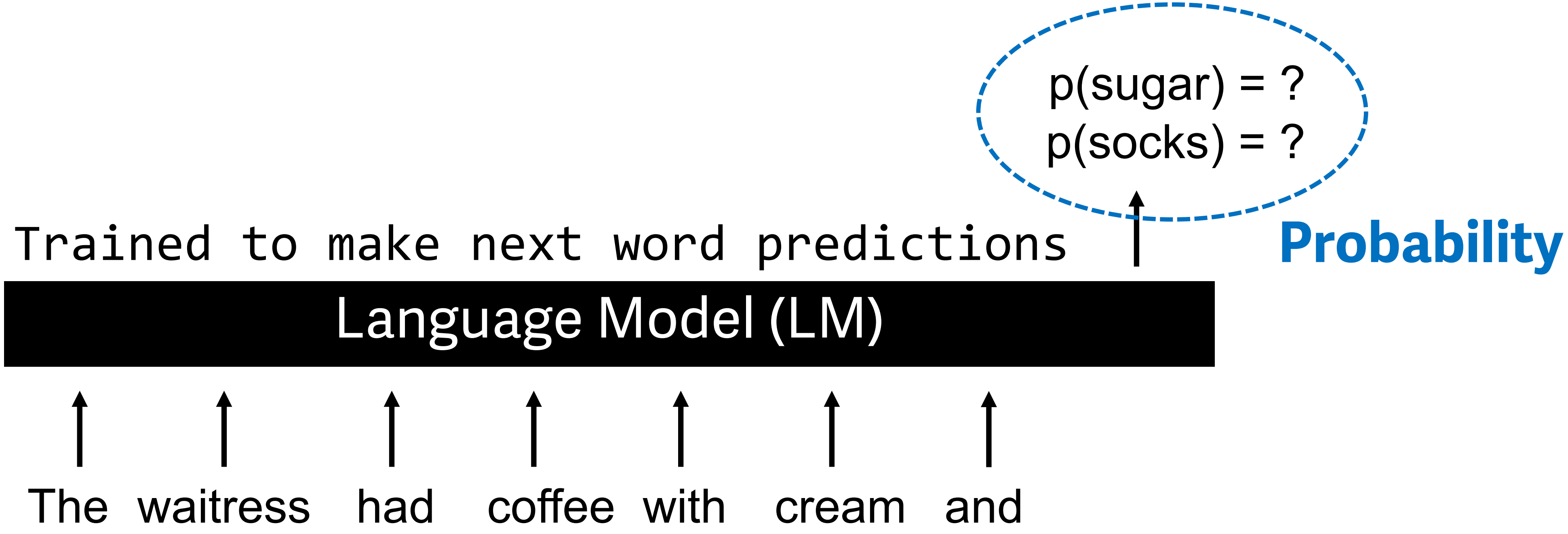
Category membership

- $\boxed{\text{palms [tree]}}$ / pines / [tree] / $\boxed{\text{tulips [flower]}}$

Unexpected; between category

Using language models for testing the hypothesis

Language models



Language models

- So far, we have discussed language models trained on..
  - Number of occurrences of words like N-gram models
  - Word predictions trained on Skip-gram and Continuous Bag-of-Words (like Word2Vec in Mikolov et al., 2013)

- Prior NN-based models had weaknesses (e.g., polysemy, out-of-vocabulary, forgetting problem, etc.)

- More recent model: Transformer-based models

Vaswani et al. (2017)

Attention Is All You Need

Ashish Vaswani* Google Brain avaswani@google.com

Noam Shazeer* Google Brain noam@google.com

Niki Parmar* Google Research nikip@google.com

Jakob Uszkoreit* Google Research usz@google.com

Llion Jones* Google Research llion@google.com

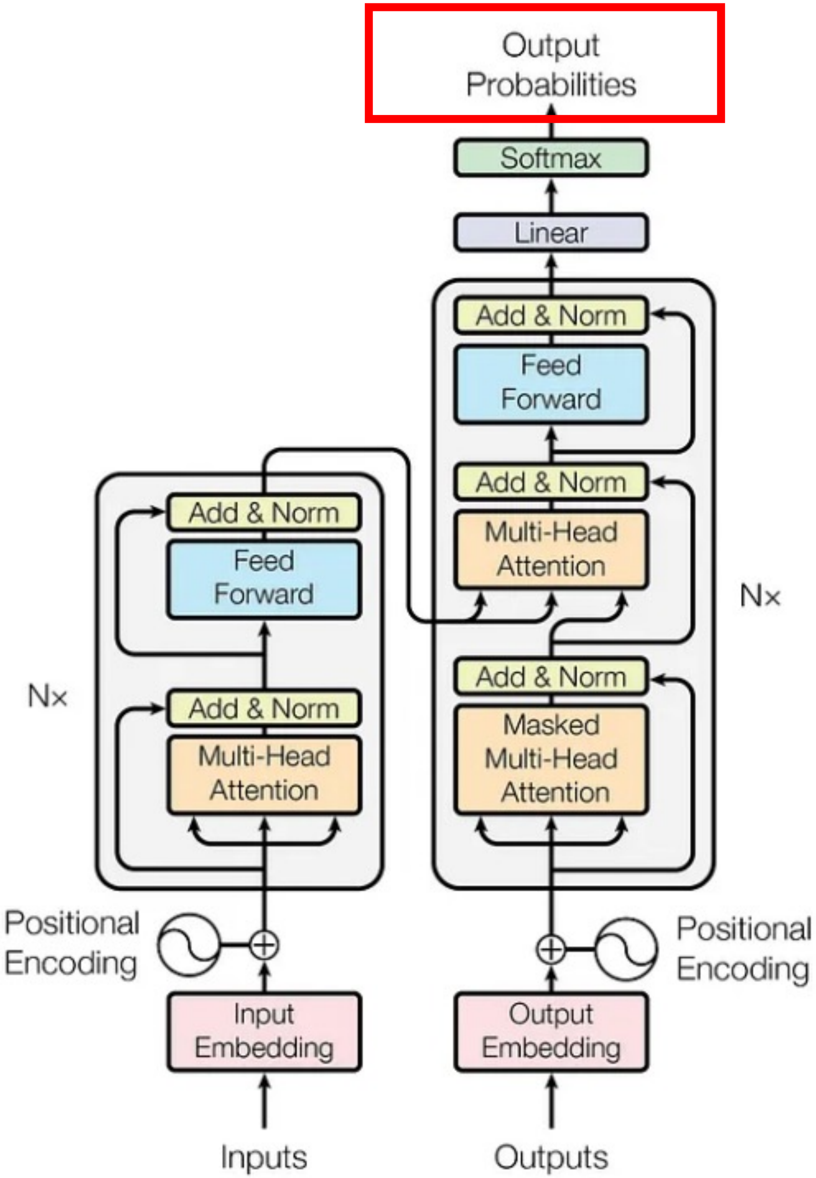
Aidan N. Gomez* † University of Toronto aidan@cs.toronto.edu

Łukasz Kaiser* Google Brain lukaszkaiser@google.com

Illia Polosukhin* ‡ illia.polosukhin@gmail.com

Abstract

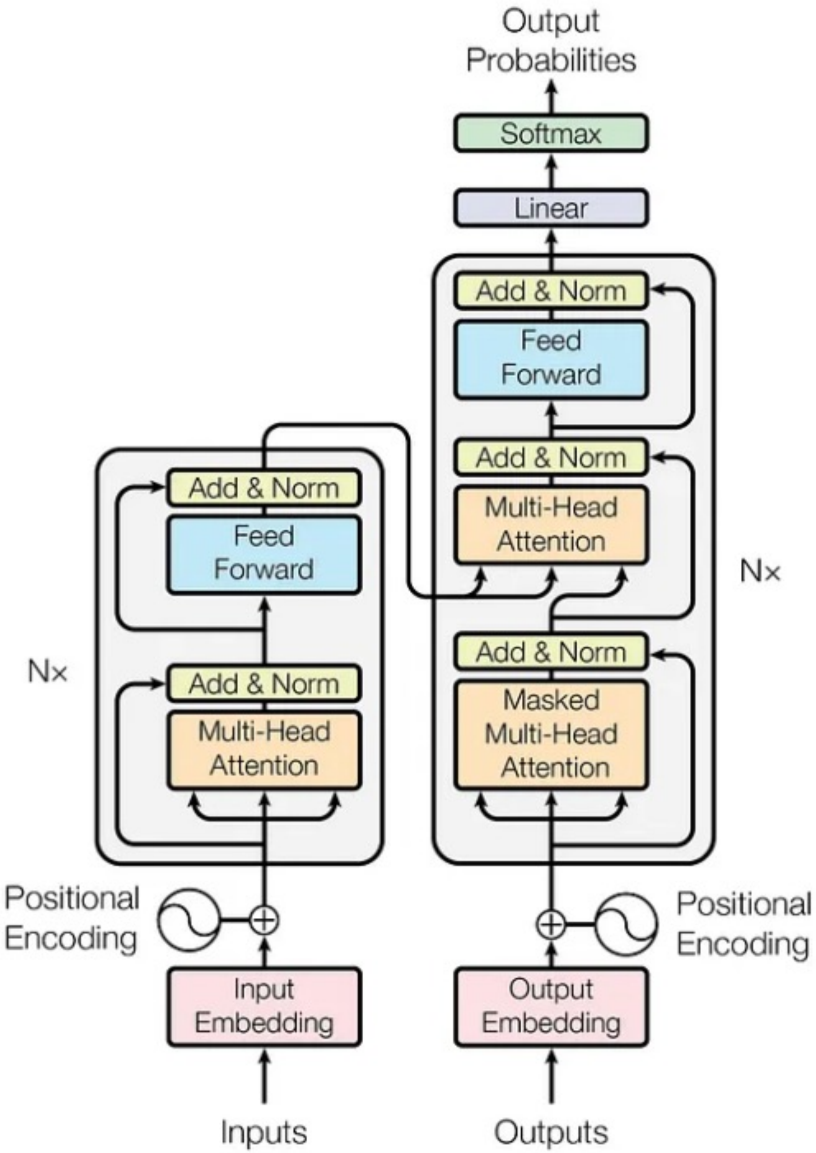
The dominant sequence transduction models are based on complex recurrent or convolutional neural networks that include an encoder and a decoder. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves 28.4 BLEU on the WMT 2014 English-to-German translation task, improving over the existing best results, including ensembles, by over 2 BLEU. On the WMT 2014 English-to-French translation task, our model establishes a new single-model state-of-the-art BLEU score of 41.0 after training for 3.5 days on eight GPUs, a small fraction of the training costs of the best models from the literature.



Transformer Architecture

BERT

Encoder



GPT

Decoder

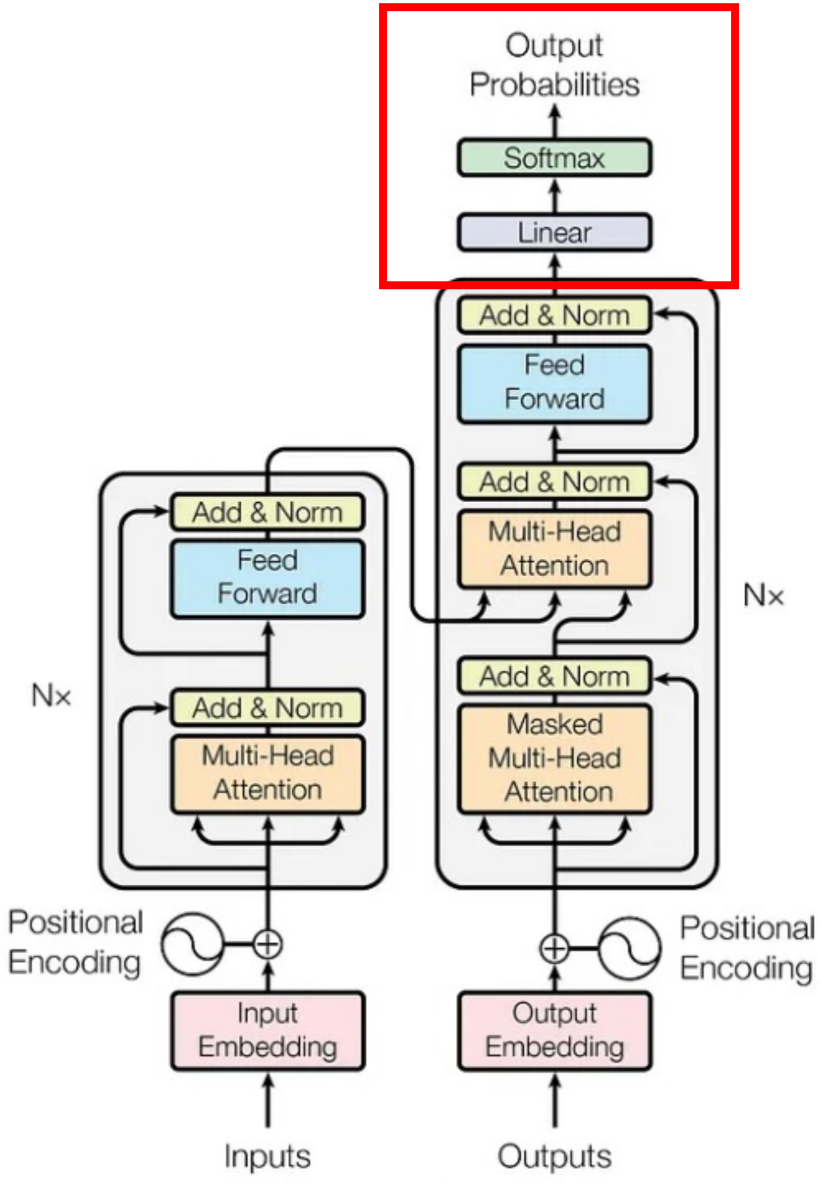
Transformer Architecture

BERT and GPT-based

- BERT-based models: Guessing the [MASK] word
  - I drank coffee with [MASK] with sugar last night.
- GPT-based models: Generating next word given prior context
  - I drank coffee with cream and ____.

The approach (similar to the assigned reading, Michaelov et al., 2023):

- Input sentences we're interested in,
- Access model output probabilities,
- Compare the model output and the observed N400 effect.



Transformer Architecture

Demo

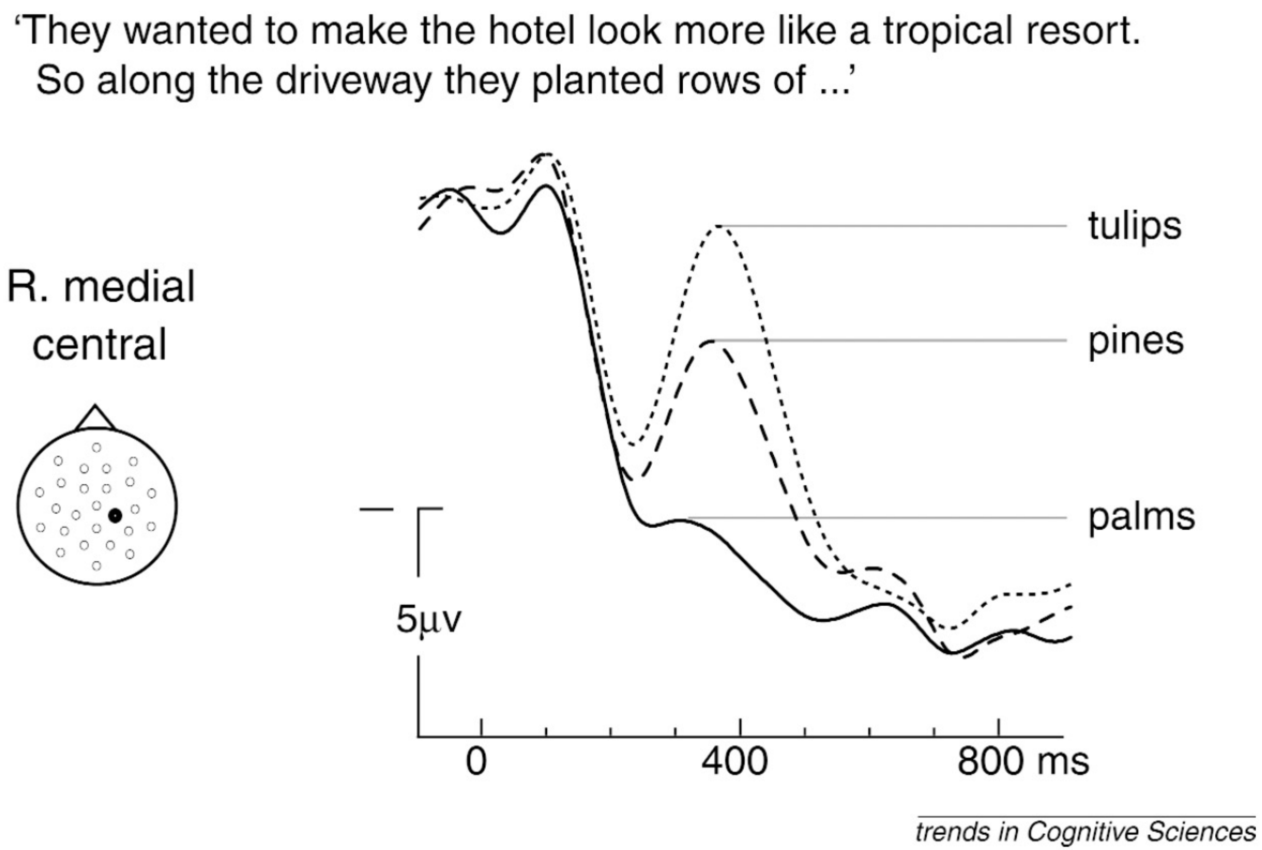
Roadmap

- Student demo (~10 min)
- Overview of Assignment 4 and FK1999 dataset (~10 min)
- R demo (~40 min)
  - Import PCIbex data
  - Analysis of Assignment 2 Demo data
  - Bar plotting results (mean and standard errors)

Assignment 4

- One-to-one relationship between cloze probability and N400 effect?
- Case study: Federmeier & Kutas (1999) report a 3-way division of the N400 effect across three conditions.

Federmeier & Kutas (1999)



Cloze probability

- palms / pines / tulips
  0.74 / < 0.05 / < 0.05

Category membership

- palms / pines / tulips
  [tree] / [tree] / [flower]

Assignment 4

- Goals:
  1. Calculate the (log) probability of the target word given context, using the FK1999 dataset and using the assigned language model.
  2. Plot bar graphs that demonstrates mean probabilities (with standard errors) based on the numbers you obtained in Step 1.
  3. Synthesize the model results with the reported findings in FK1999.

Federmeier & Kutas (1999)

	item	prefix	expected	within_category	between_category
0	1	Ann wanted to treat her foreign guests to an a...	apples	oranges	carrots
1	2	Every morning, Jack makes himself a glass of f...	oranges	apples	tomatoes
2	3	Sheila loves the taste of home-made spaghetti ...	tomatoes	carrots	apples
3	4	They told the little boy it was Bugs Bunny's f...	carrots	tomatoes	oranges
4	5	Robert saw the hive and immediately froze. He ...	bees	spiders	rats

Long form

https://pandas.pydata.org/docs/reference/api/pandas.melt.html

`pd.melt` will be useful!

	item	condition	word	prefix
0	1	expected	apples	Ann wanted to treat her foreign guests to an a...
1	1	unexpected_between	carrots	Ann wanted to treat her foreign guests to an a...
2	1	unexpected_within	oranges	Ann wanted to treat her foreign guests to an a...
3	2	expected	oranges	Every morning, Jack makes himself a glass of f...
4	2	unexpected_between	tomatoes	Every morning, Jack makes himself a glass of f...

R demo

Reading in PCIbex data and plotting a bar graph.

ggplot

https://ggplot2.tidyverse.org/articles/ggplot2.html

Week 6!