How probabilistic is language prediction in the brain? Insights from multimodal neuroimaging studies

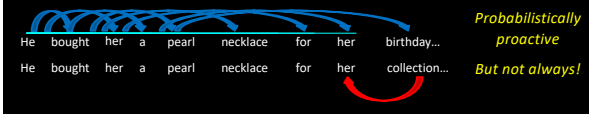
Gina Kuperberg, MD PhD

Tufts University
Athinoula A. Martinos Center for Biomedical Imaging
Mass. General Hospital

1

Language comprehension: The computational challenge

- We must be able to *proactively mobilize* our stored linguistic and non-linguistic knowledge to keep up with the pace at which the input unfolds
- We must be *flexible* enough to interpret completely unexpected inputs and adapt to new communicative situations.



Where and when?

Spatiotemporal signatures of neural responses to each incoming word in a sentence

2

This Talk

Part 1: Proactive probabilistic prediction of lexical features → early lexical facilitation

Part 2: Later retroactive processing only when we are unexpectedly surprised

Part 3: Retroactive reanalysis/re-interpretation when encounter input that *conflicts* with current communication model → Adaptation

Conclusion: Hierarchical generative framework of language comprehension.

3

This Talk

Part 1: Proactive probabilistic prediction of lexical features → early lexical facilitation

Part 2: Later retroactive processing only when we are unexpectedly surprised

Part 3: Retroactive reanalysis/re-interpretation when encounter input that *conflicts* with current communication model → Adaptation

Conclusion: Hierarchical generative framework of language comprehension.

4

Effects of contextual predictability on processing during word-by-word comprehension

Measures of word probability in relation to the preceding context:

- Cloze probability *Taylor, Journalism Quart, 1953*
- Corpus frequency *Bellegarda, Speech Commun, 2004; Jurafsky & Martin, Speech and Language Processing, 2000*
- Language Models *Devlin & Chang arXiv, 2018; Willems & Frank, Cerebral Cortex, 2015*

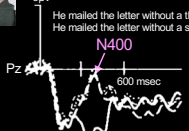
Shorter reaction times to *predictable* versus *unpredictable* words: Probabilistic effects

Lexical (or phrasal) decision
Naming
Gating
Speech monitoring
Self-paced reading *Arnon & Snider, JML, 2010; Smith & Levy, Cognition, 2013; Hintz, Meyer & Huettig, QJEP, 2016*
Eye tracking: higher skip rates and shorter fixations for more predictable words
Rayner & Well, Psychon B Rev, 1996; Rayner, Reichle, Stroud & Williams, 1996; Rayner, Li, Juhasz, & Yan, 2005; Boston & Vasishth, J Eye Movement Res, 2008; Demberg & Keller, Cognition, 2008; Kliegl & Grabner Euro J of Cogn Psych, 2004; Smith & Levy, Cognition, 2013; Luke & Christianson, Cogn Psych, 2016; Soreno & Hand, QJEP, 2017

5

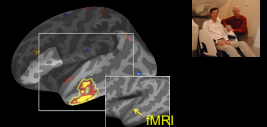
Probabilistic effects of context on neural activity to incoming words between 300-500ms

Event-related Potentials (ERPs)

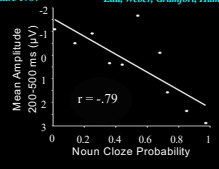


Kutas & Hillyard, Nature 1984

Magneto-encephalography (MEG)



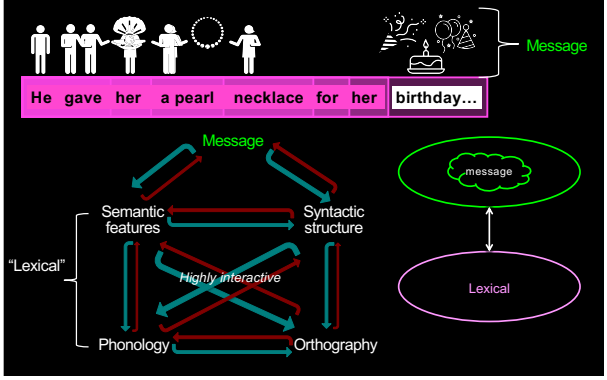
Lau, Weber, Gramfort, Hämäläinen, & Kuperberg, Cereb Cortex, 2014



DeLong, Urbach & Kutas, Nat Neurosci, 2005

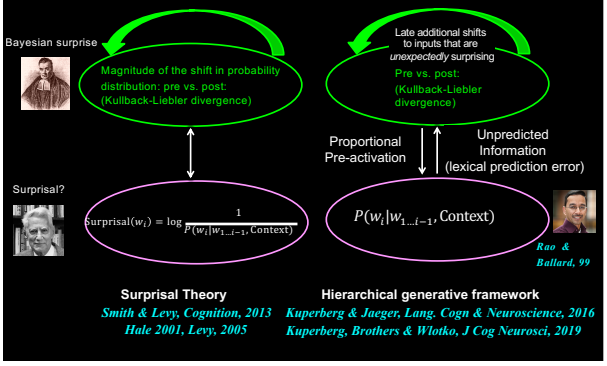
6

Incremental language comprehension



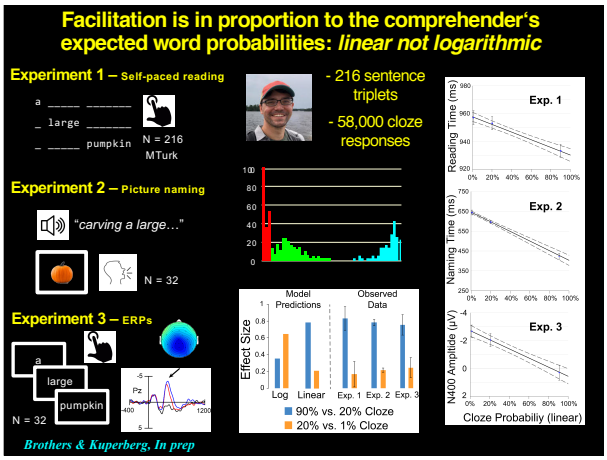
7

Probabilistic effects of context on processing each incoming word during comprehension



8

Facilitation is in proportion to the comprehender's expected word probabilities: *linear not logarithmic*

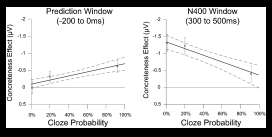


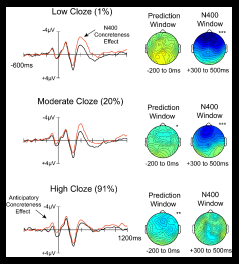
9

Experiment 3 – ERPs

Evidence for proportional pre-activation leading to proportional facilitation

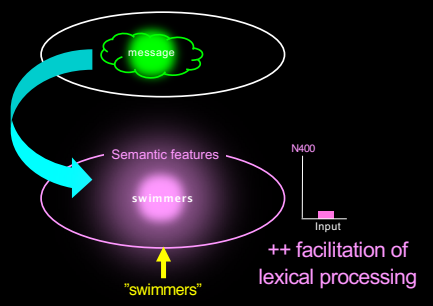
- **Concrete words** (*table*, *hand*) elicit larger N400s over frontal electrode sites than **Abstract words** (*hope*, *ideas*)
- The evoked *concreteness effect* is reduced in more predictable contexts (300-500ms)
- Instead semantic processing occurs prior to word onset





10

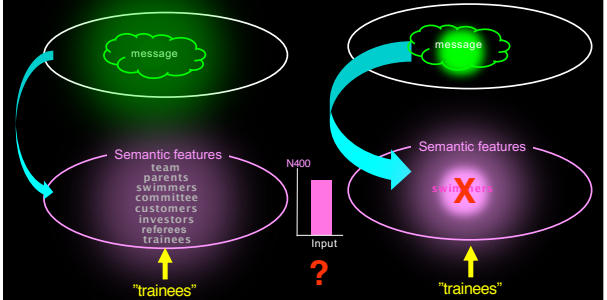
Predictable words: Easier to process because their features have already been pre-activated



11

Unpredictable words: Difficult to process because their features have not already been pre-activated

?? because costs in suppressing incorrect prior predictions

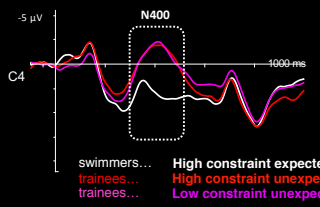


12

Unpredictable words: Difficult to process because their features have not already been pre-activated

Low constraint: *Eric and Grant received the news late in the day. They mulled over the information, and decided it was better to act sooner rather than later. Hence they cautioned the…trainees*

High constraint: *The lifeguards received a report of sharks right near the beach. Their immediate concern was to prevent any incidents in the sea. Hence they cautioned the…swimmers/trainees*



swimmers… **High constraint expected**
trainees… **High constraint unexpected**
trainees… **Low constraint unexpected**

Kutas & Hillyard, Nature 1984

Kuperberg, Brothers & Wlotko, Journal Cognitive Neuroscience, 2019

Federmeier et al., CBR, 2007

13

Unpredictable words: Difficult to process because their features have not already been pre-activated

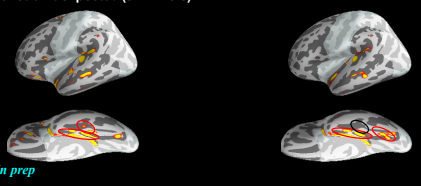
Low constraint: *Eric and Grant received the news late in the day. They mulled over the information, and decided it was better to act sooner rather than later. Hence they cautioned the…trainees*

High constraint: *The lifeguards received a report of sharks right near the beach. Their immediate concern was to prevent any incidents in the sea. Hence they cautioned the…swimmers/trainees*

MEG Source Localization: (300-500ms)

Low constraint unexpected (trainees) *minus* **High constraint expected (swimmers)**

High constraint unexpected (trainees) *minus* **High constraint expected (swimmer)**



Wang et al., In prep

14

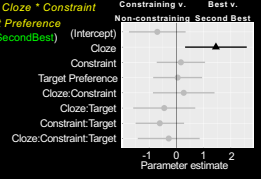
Unpredictable words: Difficult to process because their features have not already been pre-activated

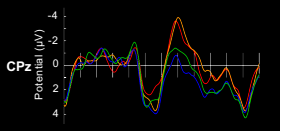
True across the whole range of probability?

Medium Constraining (span: 36-92%) Stephen wanted to do something special for his girlfriend. He decided to make her a hand-made card. On it, he drew some… **hearts** (57%) / **flowers** (16%) …and folded the paper.
(Best / SecondBest)

Non-constraining (control) Malcom always doodled in class. He took out a fresh sheet of paper. On it, he drew some…**hearts** (4%) / **flowers** (3%) …and folded the paper.
(Unexpected / Unexpected)

Linear mixed effects regression
*N400 ~ Cloze * Constraint*
** Target Preference*
(Best / SecondBest)

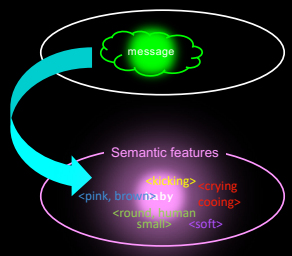


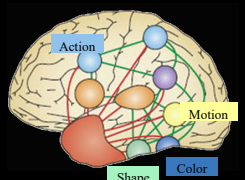


N400 amplitude driven only by cloze probability

15

Can we detect neural pre-activation of semantic features prior to encountering an incoming word?



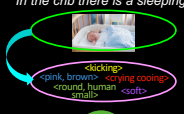


Patterson, Nestor & Rogers, Nature Reviews Neuroscience 2007

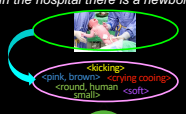
16

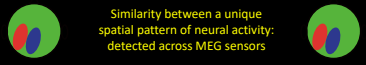
Neural evidence for *representationally-specific* lexico-semantic pre-activation

In the crib there is a sleeping…



In the hospital there is a newborn…

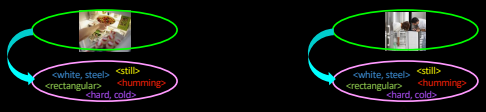


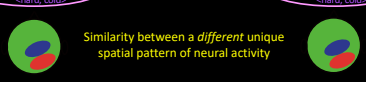


Similarity between a unique spatial pattern of neural activity: detected across MEG sensors

In order to prevent the milk from going bad, Mom put it in the …

In order to keep the food fresh, the family bought a new …





Similarity between a *different* unique spatial pattern of neural activity

17

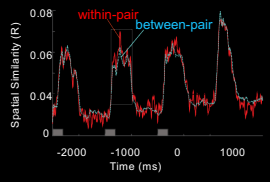
Neural evidence for *representationally-specific* lexico-semantic pre-activation

In the crib there is a sleeping …
In the hospital there is a newborn …

In order to prevent the milk from going bad, Mom put it in the …
In order to keep the food fresh, the family bought a new …

Spatial patterns of neural activity are more similar to each other when predicting the same words vs different words

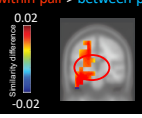
within > between:
$t_{(25)} = 4.434$, $p < .0001$



Spatial similarity at every time point

Temporal patterns of neural activity are more similar to each other when predicting the same words vs different words

within-pair > between-pair

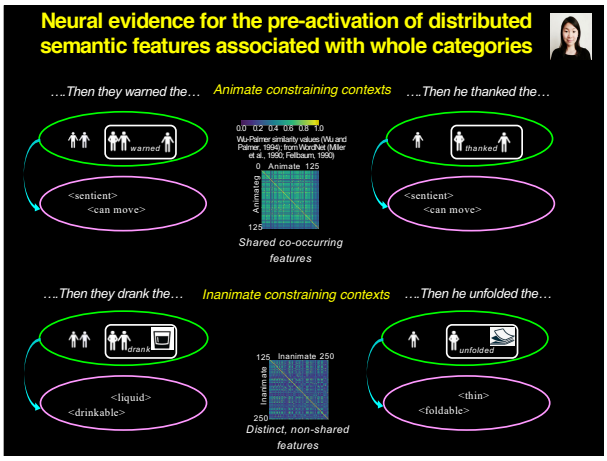


y = -21.5
p = .002

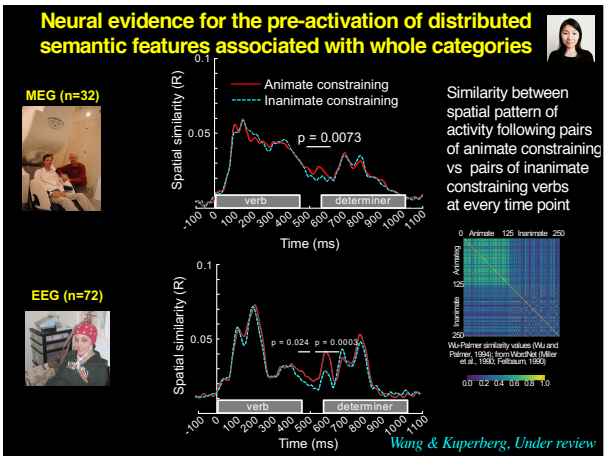
Temporal similarity at every voxel

Wang, Kuperberg & Jensen, E-life 2019

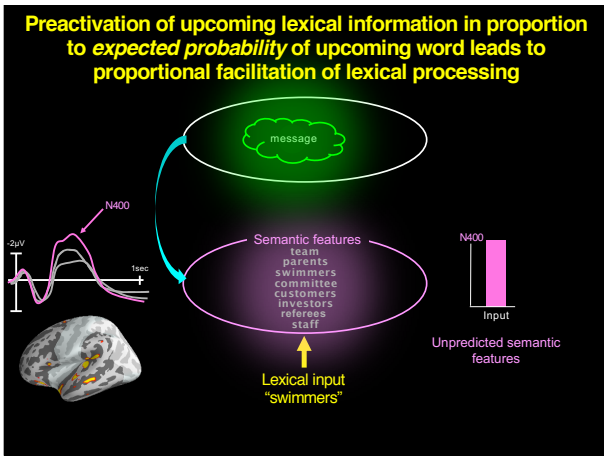
18



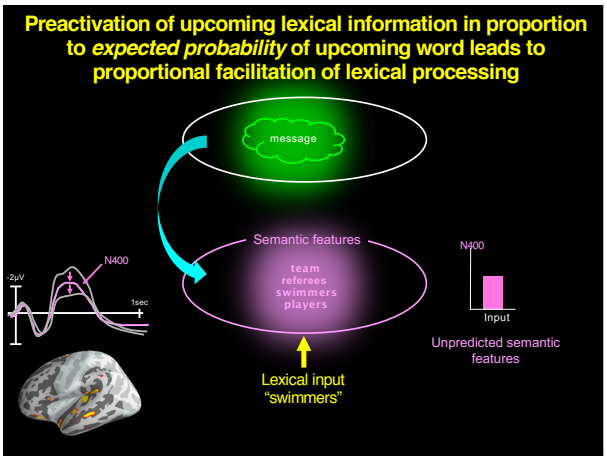
19



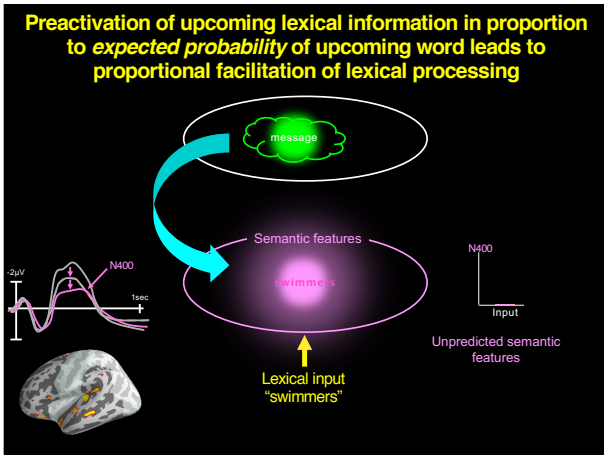
20



21



22



23

This Talk

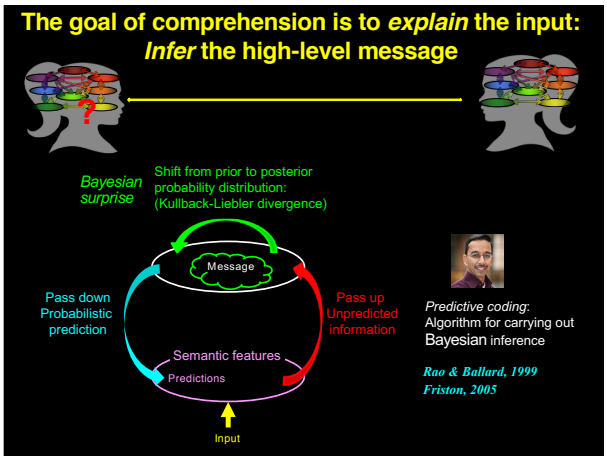
Part 1: Proactive probabilistic prediction of lexical features → early lexical facilitation

Part 2: Later retroactive processing only when we are unexpectedly surprised

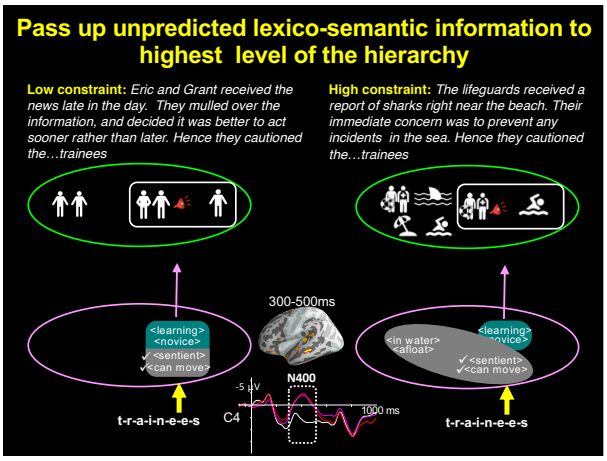
Part 3: Retroactive reanalysis/re-interpretation when encounter input that *conflicts* with current communication model → Adaptation

Conclusion: Hierarchical generative framework of language comprehension.

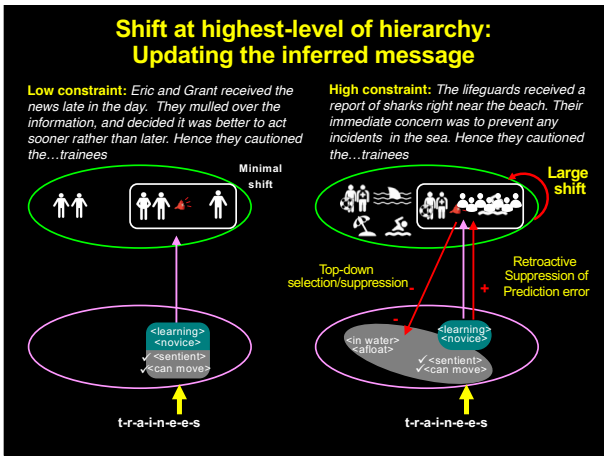
24



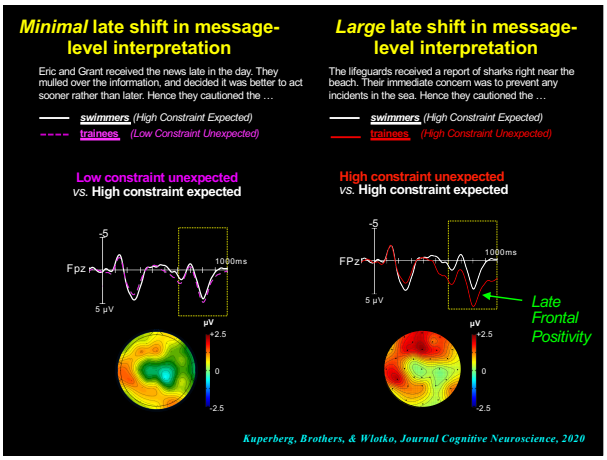
25



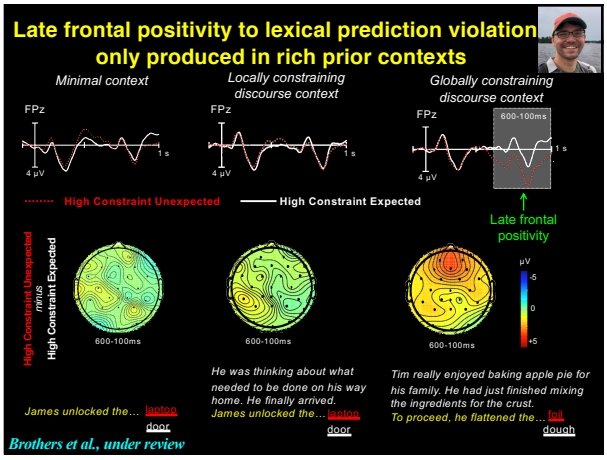
26



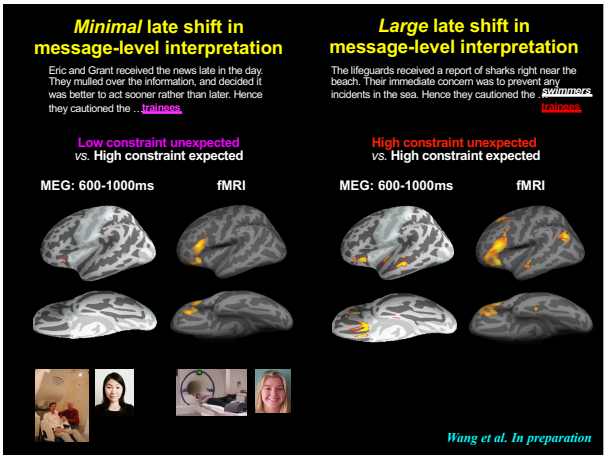
27



28



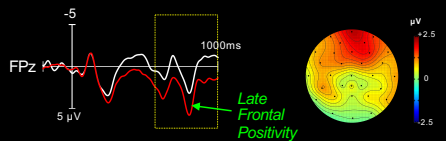
29



30

Late Frontal Positivity can also be produced when new input leads to large shift in interpretation even in low constraint contexts

Low constraint informative
vs. Low constraint uninformative



The way things were going, no one expected it to happen.
It left all of the onlookers completely speechless.
After the…

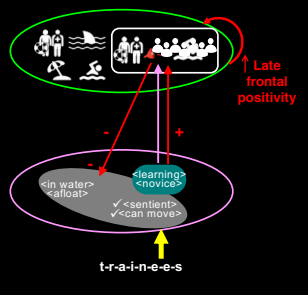
commotion *(Low Constraint Uninformative)*
touchdown *(Low Constraint Informative)*

Brothers, Kuperberg, in prep

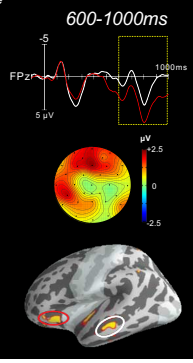
31

Summary: When unpredicted input triggers large update in message-level interpretation

The lifeguards received a report of sharks right near the beach. Their immediate concern was to prevent any incidents in the sea. Hence they cautioned the…



600-1000ms



32

This Talk

Part 1: Proactive probabilistic prediction of lexical features → early lexical facilitation

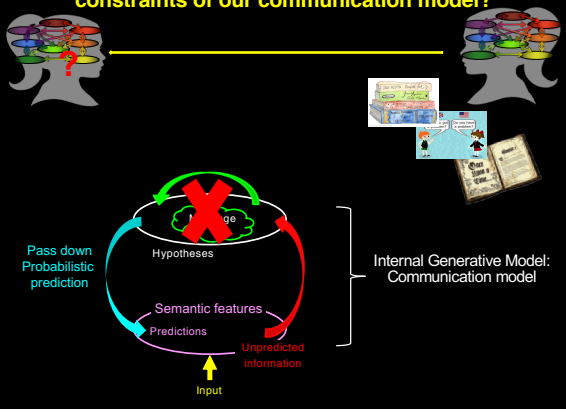
Part 2: Later retroactive processing only when we are unexpectedly surprised

Part 3: Retroactive reanalysis/re-interpretation when encounter input that *conflicts* with current communication model → Adaptation

Conclusion: Hierarchical generative framework of language comprehension.

33

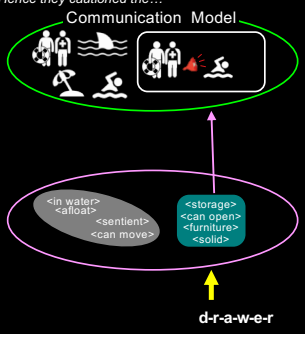
What happens when the input conflicts with the constraints of our communication model?



34

What happens when the input conflicts with the constraints of our communication model?

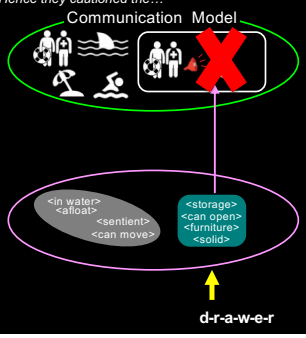
The lifeguards received a report of sharks right near the beach. Their immediate concern was to prevent any incidents in the sea. Hence they cautioned the…



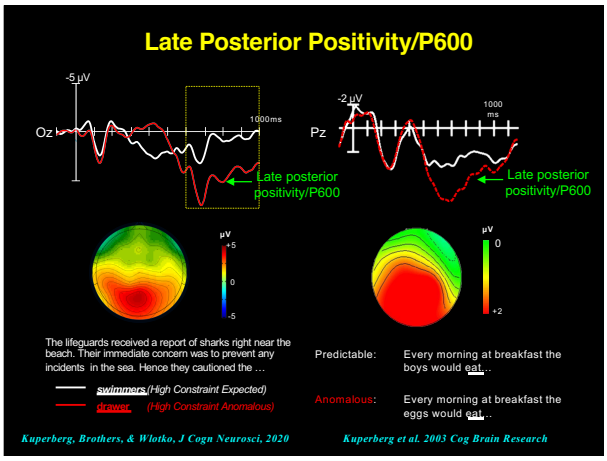
35

What happens when the input conflicts with the constraints of our communication model?

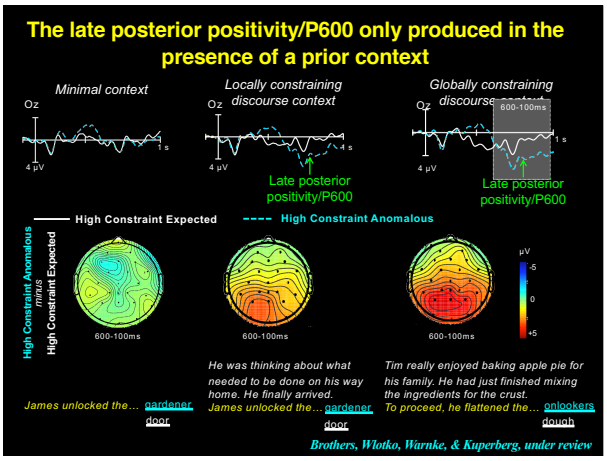
The lifeguards received a report of sharks right near the beach. Their immediate concern was to prevent any incidents in the sea. Hence they cautioned the…



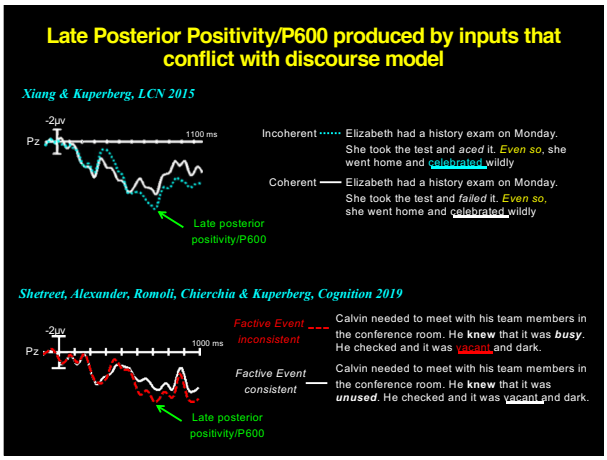
36



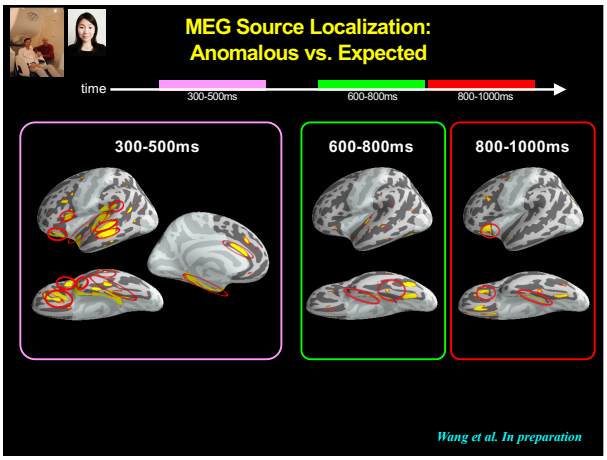
37



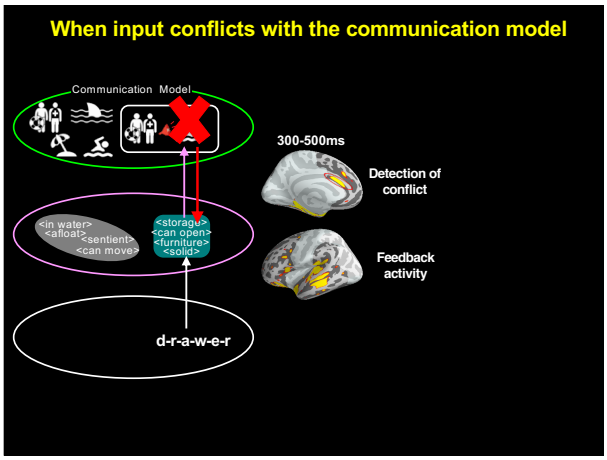
38



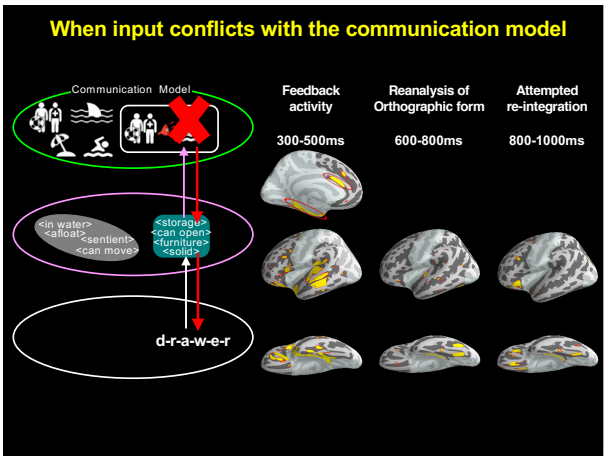
39



40



41

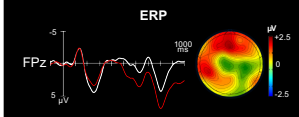


42

Summary: 600-1000ms

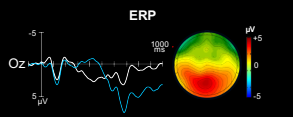
Large late shift in interpretation

High Constraint Unexpected *vs.* Expected



Reanalysis & ?reinterpretation

Anomalous *vs.* Expected



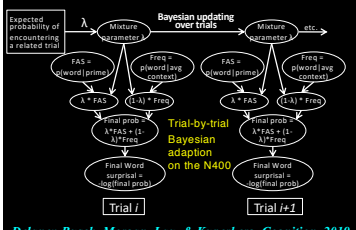
Kuperberg, Brothers & Wlotko, 2019
Wang et al. In preparation

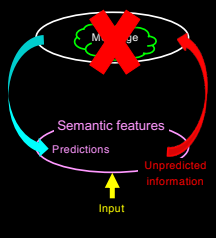
43

When inputs conflict with communication model: Trigger for fast adaptation to the broader statistical structure of our ever-changing environment?

Our communicative environments keep changing…

….fast & often in systematic ways.



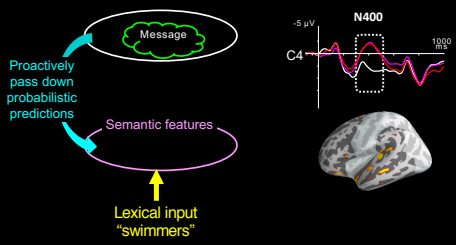


Delaney-Busch, Morgan, Lau, & Kuperberg, Cognition, 2019

44

Conclusions: Hierarchical Generative Framework of Language Comprehension

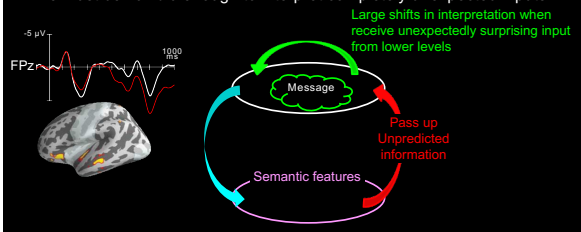
- We must be able to *proactively mobilize* our stored linguistic and non-linguistic knowledge to keep up with the pace at which the input unfolds.



45

Conclusions: Hierarchical Generative Framework of Language Comprehension

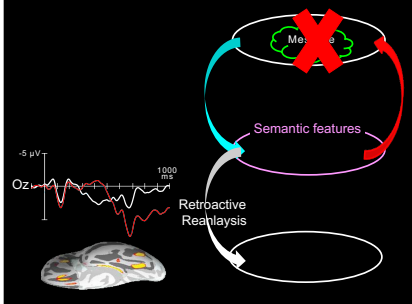
- We must be *flexible* enough to interpret completely unexpected inputs



46

Conclusions: Hierarchical Generative Framework of Language Comprehension

- We must be *flexible* enough to interpret completely anomalous inputs, try again & adapt to new communicative situations.



47

Thank you!

National Institute of Mental Health: R01MH071635
National Institute of Child Health and Human Development: R01HD082527

48