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

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.

This Talk

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

Measures of word probability in relation to the preceding context:
- Cloze probability
- Corpus frequency
- Language Models

Shorter reaction times to predictable versus unpredictable words: Probabilistic effects

Lexical (or phrasal) decision
Naming
Gating
Speech monitoring
Self-paced reading
Eye tracking: higher skip rates and shorter fixations for more predictable words

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

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.

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.
**Linear mixed effects regression**

**Target Preference**

\[ \text{N400} \sim \text{Cloze} \times \text{Constraint} \]

Best (control) constraining

Medium

Low constraint:

- sooner rather than later. Hence they cautioned the... (trainees)

High constraint:

- the... (swimmers)

**Linear mixed effects regression**

\[ \text{CPz} = \beta_0 + \beta_1 \times \text{Target Preference} + \beta_2 \times \text{Constraint} + \beta_3 \times \text{Target Preference} \times \text{Constraint} + \varepsilon \]

Parameter estimates

N400 amplitude driven only by cloze probability

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

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

In the crib there is a sleeping...

- Similarity between a unique spatial pattern of neural activation detected across MEG sensors

In the hospital there is a newborn...

- Similarity between a different unique spatial pattern of neural activity

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 ...

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

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

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

\[ \text{within-pair} > \text{between-pair} \]

**Temporal similarity at every point**

[Graph showing the results of the comparisons between different pairs of conditions.]
Neural evidence for the pre-activation of distributed semantic features associated with whole categories

Preactivation of upcoming lexical information in proportion to expected probability of upcoming word leads to proportional facilitation of lexical processing

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.
The goal of comprehension is to explain the input: infer the high-level message

Pass up unpredicted lexico-semantic information to highest level of the hierarchy

Pass up signalized information

Pass down signalized prediction

Shift from prior to posterior probability distribution (Kullback-Leibler divergence)

Predictive coding Algorithm for carrying out Bayesian inference

Bayesian surprise

Shift at highest-level of hierarchy: Updating the inferred message

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...trainees

Minimal late shift in message-level interpretation

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...trainees

Large late shift in message-level interpretation

Minimal late shift in message-level interpretation

Large late shift in message-level interpretation

Late frontal positivity to lexical prediction violation only produced in rich prior contexts

Wang et al. In preparation
Late Frontal Positivity can also be produced when new input leads to large shift in interpretation even in low constraint contexts

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

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.
**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.

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

- 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, try again & adapt to new communicative situations.

Thank you!