Prediction and memory in human language comprehension: Evidence from naturalistic fMRI
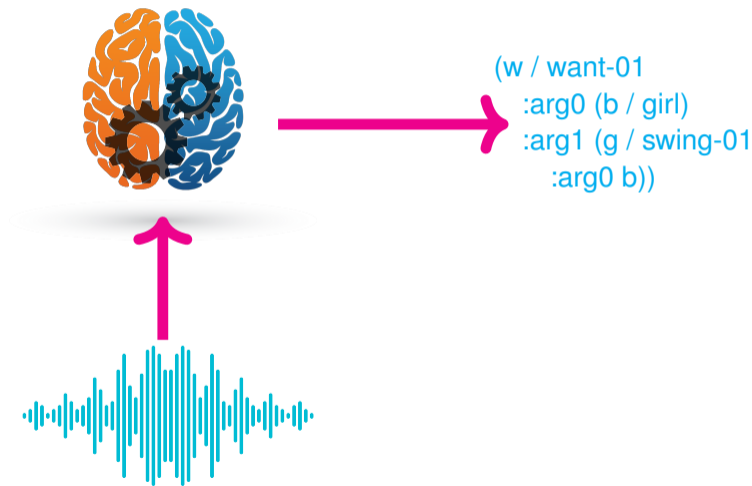
Cory Shain
(with Idan Blank, Marten van Schijndel, Evelina Fedorenko, Edward Gibson, and William Schuler)

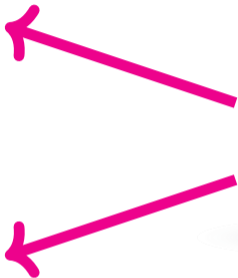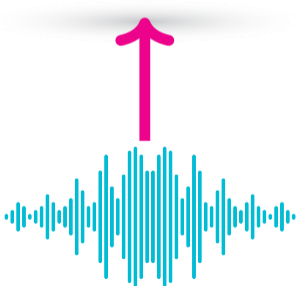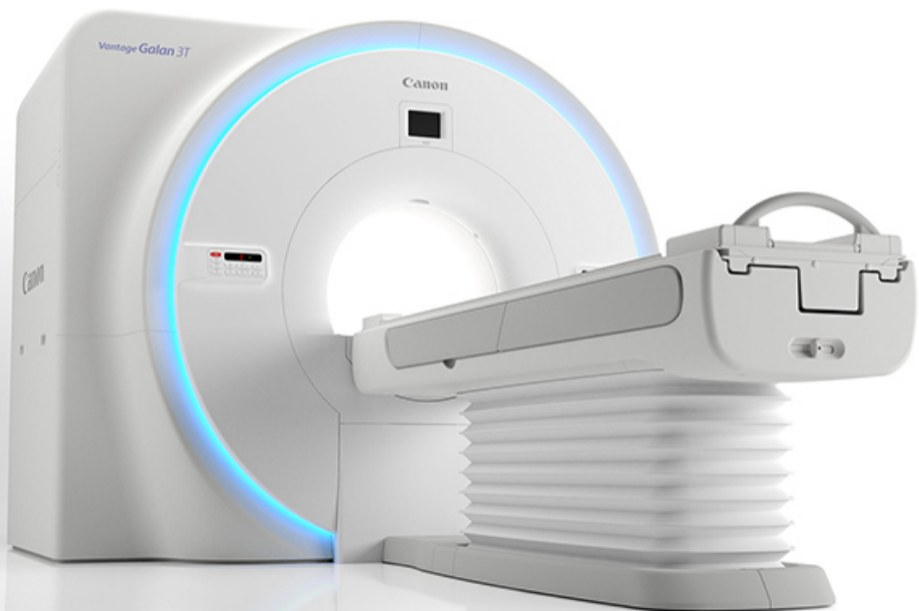Dec 7, 2020, Centre for Language Studies, Radboud University

```
(w / want-01
  :arg0 (b / girl)
  :arg1 (g / swing-01
    :arg0 b))
```

$$\int \beta e^{-\beta x} \, dx$$

```
(w / want-01
  :arg0 (b / girl)
  :arg1 (g / swing-01
    :arg0 b))
```
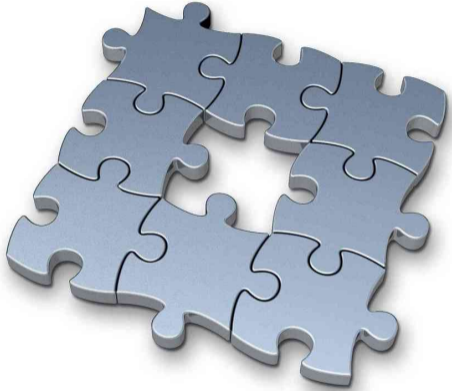
+ Prediction ⊂ comprehension
  (cf. Levy 2008)

+ There are dedicated working memory and predictive coding resources for language
  (cf. Fedorenko et al. 2006; Huettig and Mani 2016)

+ Syntactic analysis is a core subroutine in typical language processing
  (cf. Swets et al. 2008; Frank and Bod 2011)

+ Some costs don't register in reading

+ Prediction $\subset$ comprehension
  (cf. Levy 2008)
+ There are dedicated working memory and predictive coding resources for language
  (cf. Fedorenko et al. 2006; Huettig and Mani 2016)
+ Syntactic analysis is a core subroutine in typical language processing
  (cf. Swets et al. 2008; Frank and Bod 2011)
+ Some costs don't register in reading

+ Prediction $\subset$ comprehension
  (cf. Levy 2008)

+ There are dedicated working memory and predictive coding resources for language
  (cf. Fedorenko et al. 2006; Huettig and Mani 2016)

+ Syntactic analysis is a core subroutine in typical language processing
  (cf. Swets et al. 2008; Frank and Bod 2011)

+ Some costs don't register in reading

+ Prediction ⊂ comprehension
  (cf. Levy 2008)
+ There are dedicated working memory and predictive coding resources for language
  (cf. Fedorenko et al. 2006; Huettig and Mani 2016)
+ Syntactic analysis is a core subroutine in typical language processing
  (cf. Swets et al. 2008; Frank and Bod 2011)
+ Some costs don't register in reading

+ Memory and prediction effects are central to psycholinguistics
(Grodner and Gibson 2005; Levy 2008; Kuperberg and Jaeger 2016; Ferreira and Chantavarin 2018)

+ Can differentiate:

+ Both functions are domain general

+ Memory and prediction effects are central to psycholinguistics
  (Grodner and Gibson 2005; Levy 2008; Kuperberg and Jaeger 2016; Ferreira and Chantavarin 2018)
+ Can differentiate:
  + Processing strategies
    (Grodner and Gibson 2005; Lewis and Vasishth 2005; Hale 2001; Levy 2008)
  + Representations
    (Brennan et al. 2016; Lopopolo et al. 2020)
+ Both functions are domain general
  (Federmeier et al. 2020)

Memory and Prediction in Sentence Processing

+ Memory and prediction effects are central to psycholinguistics
  (Grodner and Gibson 2005; Levy 2008; Kuperberg and Jaeger 2016; Ferreira and Chantavarin 2018)
+ Can differentiate:
  + Processing strategies
    (Grodner and Gibson 2005; Lewis and Vasishth 2005; Hale 2001; Levy 2008)
  + Representations
    (Brennan et al. 2016; Lopopolo et al. 2020)
+ Both functions are domain general
  (Federmeier et al. 2020)

Memory and Prediction in Sentence Processing

+ Memory and prediction effects are central to psycholinguistics
  (Grodner and Gibson 2005; Levy 2008; Kuperberg and Jaeger 2016; Ferreira and Chantavarin 2018)
+ Can differentiate:
  + Processing strategies
    (Grodner and Gibson 2005; Lewis and Vasishth 2005; Hale 2001; Levy 2008)
  + Representations
    (Brennan et al. 2016; Lopopolo et al. 2020)
+ Both functions are domain general
  (Federmeier et al. 2020)

Memory and Prediction in Sentence Processing

+ Memory and prediction effects are central to psycholinguistics
  (Grodner and Gibson 2005; Levy 2008; Kuperberg and Jaeger 2016; Ferreira and Chantavarin 2018)
+ Can differentiate:
    + Processing strategies
      (Grodner and Gibson 2005; Lewis and Vasishth 2005; Hale 2001; Levy 2008)
    + Representations
      (Brennan et al. 2016; Lopopolo et al. 2020)
+ Both functions are domain general
  (Federmeier et al. 2020)

How to investigate?

How to investigate?

Shared resources? → **imaging**

How to investigate?

Shared resources? → imaging
Observer's paradox → **naturalistic stimuli**

+ **These studies:** Natural Stories corpus, audio, 78 fMRI participants
  (Futrell et al. 2018)

+ Naturalistic stimuli mitigate observer's paradox
  (Hasson and Honey 2012)

+ **These studies:** Natural Stories corpus, audio, 78 fMRI participants
  (Futrell et al. 2018)

+ Naturalistic stimuli mitigate observer's paradox
  (Hasson and Honey 2012)

  + Constructed manipulations may engage other processing strategies
    (Campbell and Tyler 2018; Hasson et al. 2018; Diachek et al. 2019)
  + Effects have failed to generalize to naturalistic stim
    E.g. Grodner and Gibson 2005 vs. van Schijndel and Schuler 2013

+ **These studies:** Natural Stories corpus, audio, 78 fMRI participants
  (Futrell et al. 2018)

+ Naturalistic stimuli mitigate observer's paradox
  (Hasson and Honey 2012)

   + Constructed manipulations may engage other processing strategies
     (Campbell and Tyler 2018; Hasson et al. 2018; Diachek et al. 2019)

      + Effects have failed to generalize to naturalistic stim
        E.g. Grodner and Gibson 2005 vs. van Schijndel and Schuler 2013

+ **These studies:** Natural Stories corpus, audio, 78 fMRI participants
  (Futrell et al. 2018)

+ Naturalistic stimuli mitigate observer's paradox
  (Hasson and Honey 2012)

  + Constructed manipulations may engage other processing strategies
    (Campbell and Tyler 2018; Hasson et al. 2018; Diachek et al. 2019)
  + Effects have failed to generalize to naturalistic stim
    E.g. Grodner and Gibson 2005 vs. van Schijndel and Schuler 2013

How to investigate?

Shared resources? → imaging
Observer's paradox → **naturalistic stimuli**

How to investigate?

Shared resources? → imaging
Observer's paradox → naturalistic stimuli
Variable functional anatomy → **functional localization**

+ Two coherent functional networks:
    + Fronto-temporal language network (LANG)
    (Fedorenko et al. 2010)
    + Fronto-parietal multiple-demand network (MD)
    (Duncan 2010)

+ Two coherent functional networks:
    + Fronto-temporal language network (LANG)
      (Fedorenko et al. 2010)
    + Fronto-parietal multiple-demand network (MD)
      (Duncan 2010)

+ Two coherent functional networks:
    + Fronto-temporal language network (LANG)
      (Fedorenko et al. 2010)
    + Fronto-parietal multiple-demand network (MD)
      (Duncan 2010)
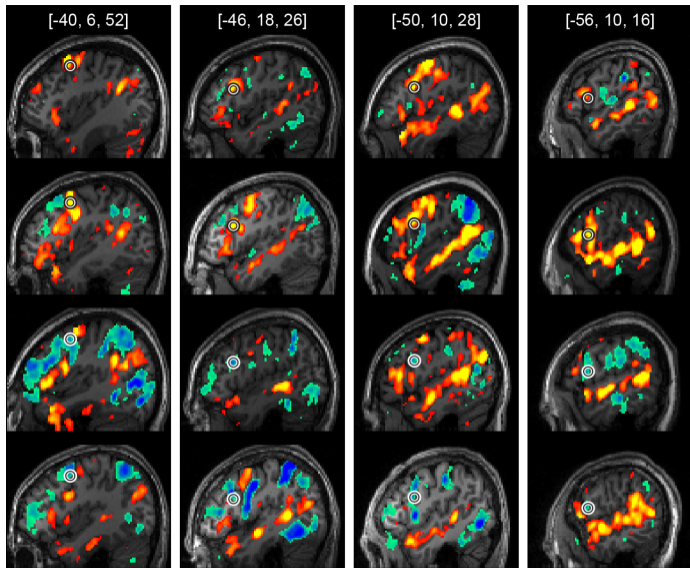
A RUSTY LOCK WAS FOUND IN THE DRAWER

A RUSTY LOCK WAS FOUND IN THE DRAWER

DAP DRELLO SMOP UB PLID KAV CRE REPLODE

[-40, 6, 52]   [-46, 18, 26]   [-50, 10, 28]   [-56, 10, 16]

Nonwords > Sentences

10^{-15} 10^{-12} 10^{-9} 10^{-6} 0.005
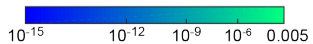
Sentences > Nonwords

0.005 10^{-6} 10^{-9} 10^{-12} 10^{-15}

How to investigate?

Shared resources? → imaging
Observer's paradox → naturalistic stimuli
Variable functional anatomy → **functional localization**

How to investigate?

Shared resources?	→	imaging
Observer's paradox	→	naturalistic stimuli
Variable functional anatomy	→	functional localization
Variable hemodynamics	→	**deconvolutional regression**
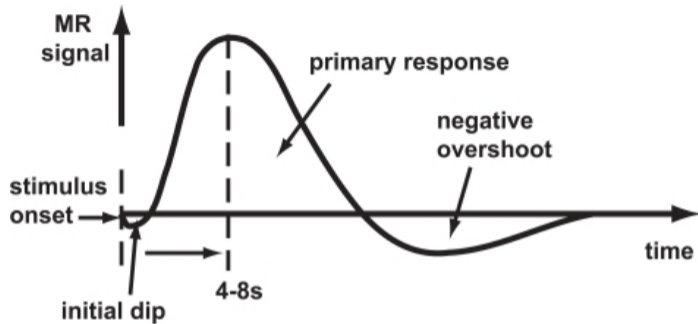
Variable Dynamics
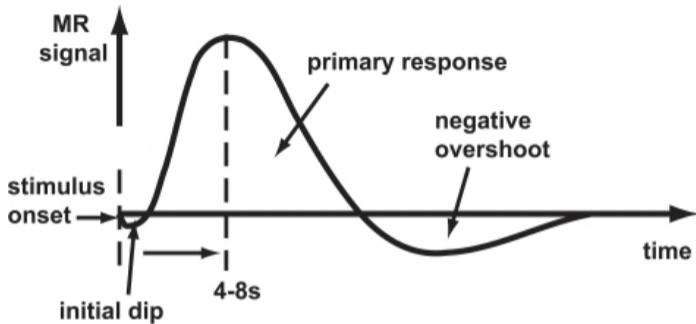
Varies by individual/region

(Handwerker et al. 2004)
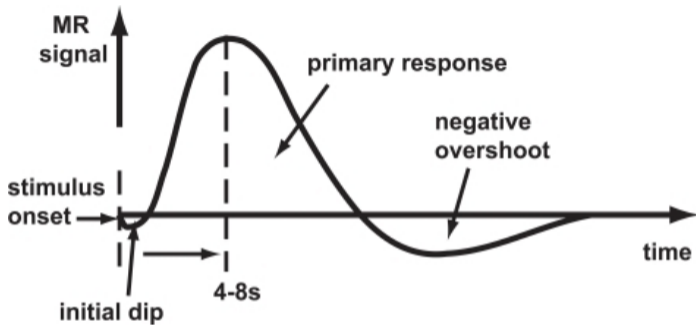
Varies by individual/region

(Handwerker et al. 2004)

Estimate HRF using continuous-time deconvolutional regression (CDR)

(Shain and Schuler 2018; Shain and Schuler 2019)

Prediction

Q1: Does syntax inform word predictions?

Q1: Does syntax inform word predictions?

Q2: Does word prediction recruit domain-general resources?

+ Prior studies report prediction effects mostly in language (LANG) regions
  (Willems et al. 2015; Brennan et al. 2016; Henderson et al. 2016; Lopopolo et al. 2017)

+ Didn't localize MD to control for variation in functional brain anatomy
  (Poldrack 2006; Fedorenko et al. 2010; Frost and Goebel 2012)

How specialized is predictive language processing?

+ Prior studies report prediction effects mostly in language (LANG) regions
  (Willems et al. 2015; Brennan et al. 2016; Henderson et al. 2016; Lopopolo et al. 2017)

+ Didn't localize MD to control for variation in functional brain anatomy
  (Poldrack 2006; Fedorenko et al. 2010; Frost and Goebel 2012)

+ Critical variables:
    + **5-gram surprisal**
    + **PCFG surprisal**

+ Critical variables:
    + **5-gram surprisal**
    + PCFG surprisal

+ Critical variables:
    + **5-gram surprisal**
    + **PCFG surprisal**

Results: Prediction
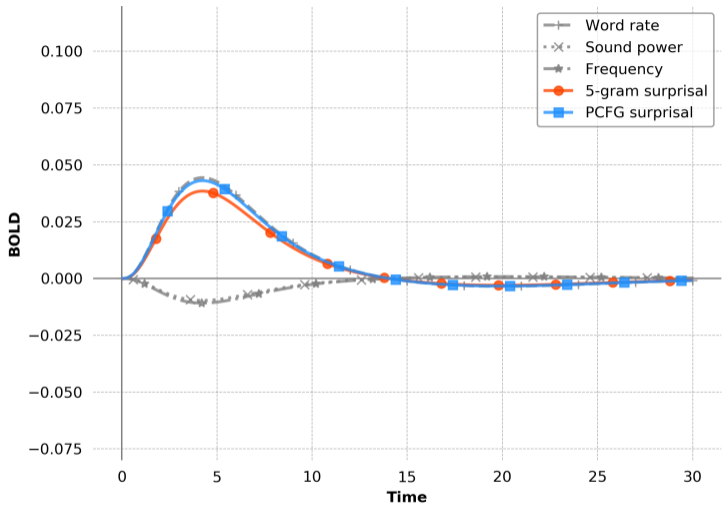
LANG network
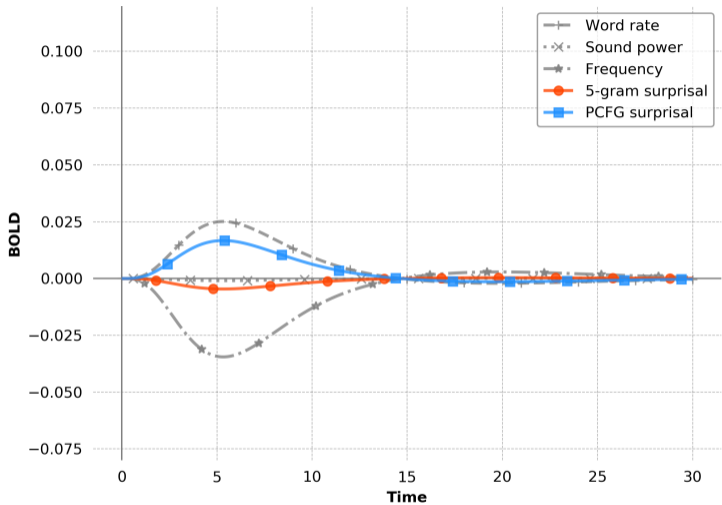
MD network
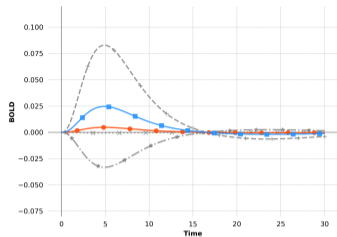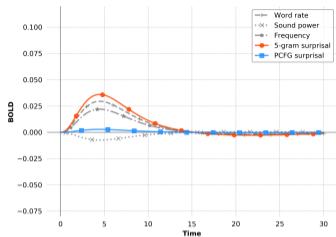
LH Angular gyrus   LH Anterior temporal lobe   LH Inferior frontal gyrus
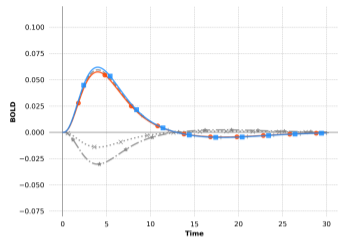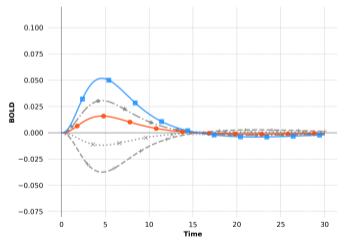
LH Inferior frontal gyrus (orbital)   LH Middle frontal gyrus   LH Posterior temporal lobe
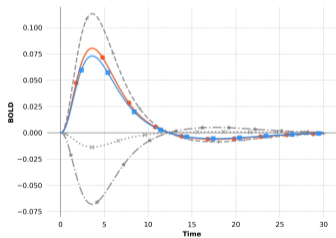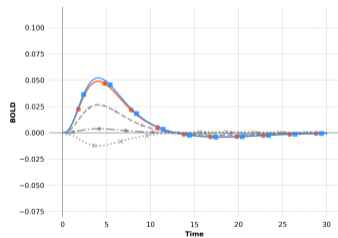
OOS hypothesis tests: LANG

Baseline model
(both effects ablated)

OOS hypothesis tests: LANG

Baseline model
(both effects ablated)

p = 0.0001***

5-gram model
(PCFG ablated)

OOS hypothesis tests: LANG

PCFG model
(5-gram ablated)

p = 0.0001***

Baseline model
(both effects ablated)

p = 0.0001***

5-gram model
(PCFG ablated)

OOS hypothesis tests: LANG

PCFG model
(5-gram ablated)

p = 0.0001***

Baseline model
(both effects ablated)

Both model
(both effects included)

p = 0.0001***

p = 0.0001***

5-gram model
(PCFG ablated)

OOS hypothesis tests: LANG

PCFG model
(5-gram ablated)

p = 0.0001***   *p* = 0.0001***

Baseline model
(both effects ablated)

Both model
(both effects included)

p = 0.0001***   *p* = 0.0001***

5-gram model
(PCFG ablated)

OOS hypothesis tests: LANG

PCFG model
(5-gram ablated)

p = 0.0001***

p = 0.0001***

Baseline model
(both effects ablated)

Both model
(both effects included)

p = 0.0001***

p = 0.0001***

5-gram model
(PCFG ablated)
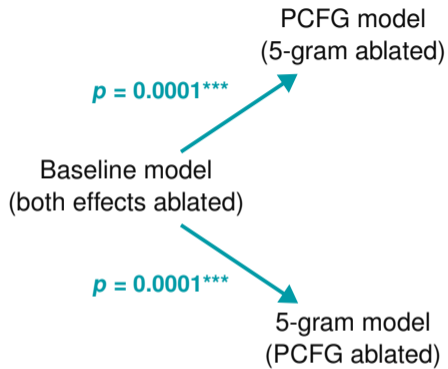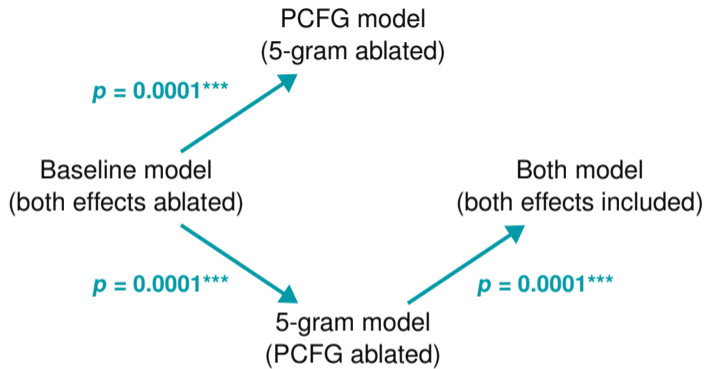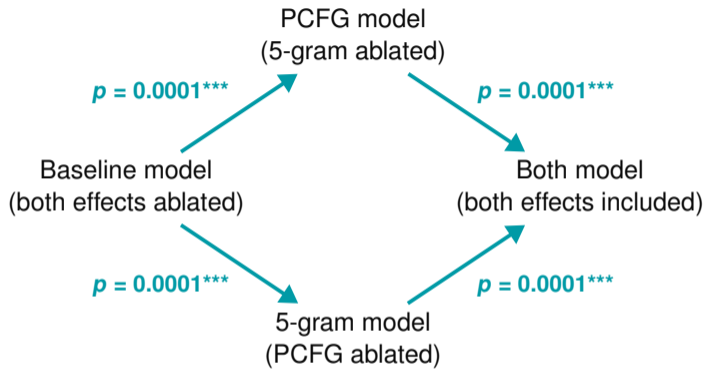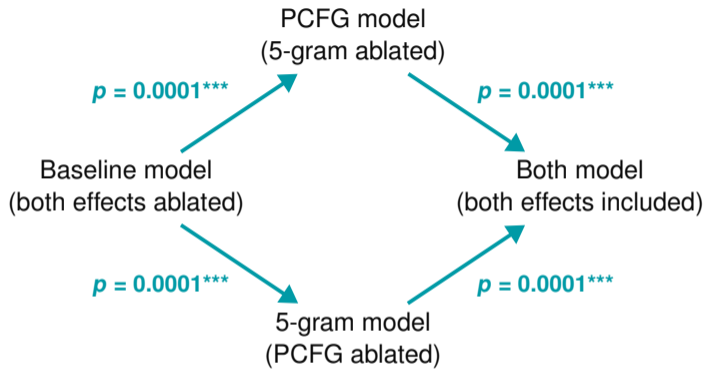
OOS hypothesis tests: LANG
Both 5-gram and PCFG surprisal explain significant OOS variance

OOS hypothesis tests: MD

Baseline model
(both effects ablated)

OOS hypothesis tests: MD

Baseline model
(both effects ablated)

$p = 0.137$

5-gram model
(PCFG ablated)

OOS hypothesis tests: MD

PCFG model
(5-gram ablated)

p = 1.000

Baseline model
(both effects ablated)

p = 0.137

5-gram model
(PCFG ablated)

OOS hypothesis tests: MD

PCFG model
(5-gram ablated)

p = 1.000

Baseline model
(both effects ablated)

Both model
(both effects included)

p = 0.137

p = 1.000

5-gram model
(PCFG ablated)
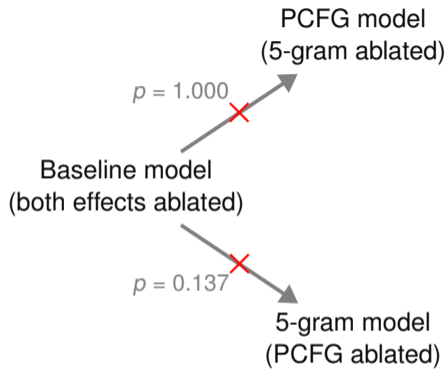
OOS hypothesis tests: MD

OOS hypothesis tests: MD

PCFG model
(5-gram ablated)

$p = 1.000$

$p = 1.000$

Baseline model
(both effects ablated)

Both model
(both effects included)

$p = 0.137$

$p = 1.000$

5-gram model
(PCFG ablated)
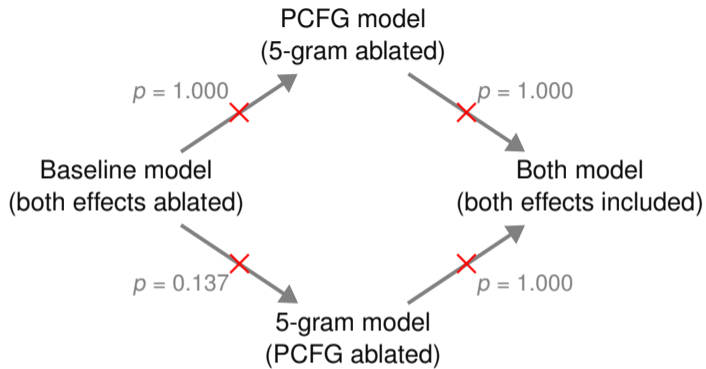
OOS hypothesis tests: MD
Neither 5-gram nor PCFG surprisal explain significant OOS variance

Q1: Does syntax inform word predictions?

Q1: Does syntax inform word predictions?
Yes

Q1: Does syntax inform word predictions?
Yes

Q2: Does word prediction recruit domain-general resources?

Q1: Does syntax inform word predictions?
Yes

Q2: Does word prediction recruit domain-general resources?
No

Q1: Does syntax inform word predictions?
Yes

Q2: Does word prediction recruit domain-general resources?
No

Shain, Blank, van Schijndel, Schuler, Fedorenko (2020). fMRI reveals language-specific predictive coding during naturalistic sentence comprehension. *Neuropsychologia*. 138.

Memory

The reporter who the senator attacked disliked the editor.

The reporter who the senator attacked **disliked** the editor.

The **reporter** who the senator attacked **disliked** the editor.

The **reporter** who the **senator** **attacked** **disliked** the editor.

subj

The reporter who the senator attacked disliked the editor.
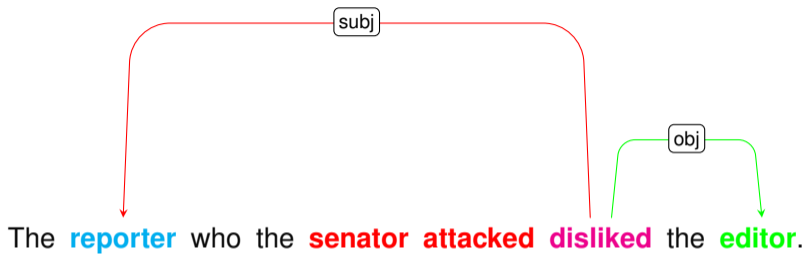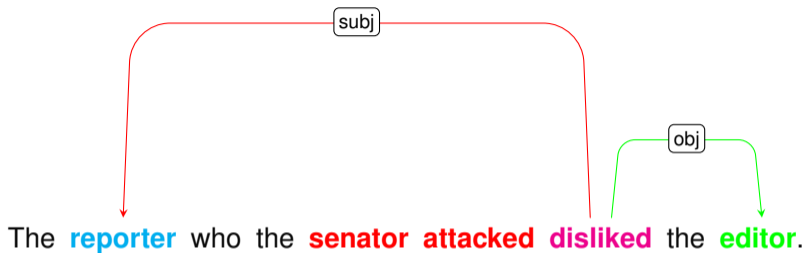
The **reporter** who the **senator** **attacked** **disliked** the **editor**.

Q3: Does language comprehension involve memory retrieval?

(Gibson 2000; Lewis and Vasishth 2005)

The **reporter** who the **senator** **attacked** **disliked** the **editor**.
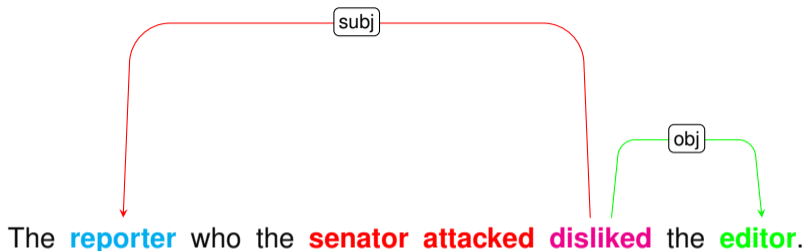
Q3: Does language comprehension involve memory retrieval?

(Gibson 2000; Lewis and Vasishth 2005)

Q4: Are memory stores are domain-general?

(Stowe et al. 1998; Fedorenko et al. 2006)

+ Retrieval effects should happen all the time
+ Retrieval effort ≠ prediction effort
  (cf. e.g. Levy 2008)

+ Retrieval effects should happen all the time
+ Retrieval effort ≠ prediction effort
  (cf. e.g. Levy 2008)

+ Prior evidence of memory effects mostly
  (Grodner and Gibson 2005)
  + Constructed/artificial stimuli
  + Little/no control for predictability

+ Memory effects null/neg using **naturalistic stimuli** with **predictability controls**
  (Demberg and Keller 2008; van Schijndel and Schuler 2013)

Memory-based comprehension

+ Prior evidence of memory effects mostly
  (Grodner and Gibson 2005)
  + Constructed/artificial stimuli
    + Little/no control for predictability
+ Memory effects null/neg using **naturalistic stimuli** with **predictability controls**
  (Demberg and Keller 2008; van Schijndel and Schuler 2013)

Memory-based comprehension

+ Prior evidence of memory effects mostly
  (Grodner and Gibson 2005)
  + Constructed/artificial stimuli
  + Little/no control for predictability
+ Memory effects null/neg using **naturalistic stimuli** with **predictability controls**
  (Demberg and Keller 2008; van Schijndel and Schuler 2013)

Memory-based comprehension

+ Prior evidence of memory effects mostly
  (Grodner and Gibson 2005)
    + Constructed/artificial stimuli
    + Little/no control for predictability
+ Memory effects null/neg using **naturalistic stimuli** with **predictability controls**
  (Demberg and Keller 2008; van Schijndel and Schuler 2013)

+ Does the language system have its own memory stores?
  + Yes
    (Caplan and Waters 1999; Fiebach et al. 2001)
  + No
    (Stowe et al. 1998; Fedorenko et al. 2006)
+ If **no**, should be memory effects in MD

+ Does the language system have its own memory stores?
    + Yes
      (Caplan and Waters 1999; Fiebach et al. 2001)
    + No
      (Stowe et al. 1998; Fedorenko et al. 2006)
+ If **no**, should be memory effects in MD

+ Does the language system have its own memory stores?
    + Yes
      (Caplan and Waters 1999; Fiebach et al. 2001)
    + No
      (Stowe et al. 1998; Fedorenko et al. 2006)
+ If **no**, should be memory effects in MD

+ Does the language system have its own memory stores?
    + Yes
        (Caplan and Waters 1999; Fiebach et al. 2001)
    + No
        (Stowe et al. 1998; Fedorenko et al. 2006)
+ If **no**, should be memory effects in MD

+ Critical variable:
    + **Dependency locality theory integration cost (DLT)**
+ Predictability controls:

+ Critical variable:
    + **Dependency locality theory integration cost (DLT)**
+ Predictability controls:
    + n-gram surprisal
    + PCFG surprisal
    + Semantic roles surprisal
    + Pre-terminal-guided surprisal

+ Critical variable:
    + **Dependency locality theory integration cost (DLT)**
+ Predictability controls:
    + 5-gram surprisal
    + PCFG surprisal
    + Adaptive RNN surprisal
        (van Schijndel and Linzen 2018)

+ Critical variable:
    + **Dependency locality theory integration cost (DLT)**
+ Predictability controls:
    + 5-gram surprisal
    + PCFG surprisal
    + Adaptive RNN surprisal
        (van Schijndel and Linzen 2018)

+ Critical variable:
    + **Dependency locality theory integration cost (DLT)**
+ Predictability controls:
    + 5-gram surprisal
    + PCFG surprisal
    + Adaptive RNN surprisal
      (van Schijndel and Linzen 2018)

+ Critical variable:
  + **Dependency locality theory integration cost (DLT)**
+ Predictability controls:
  + 5-gram surprisal
  + PCFG surprisal
  + Adaptive RNN surprisal
    (van Schijndel and Linzen 2018)

Results: Memory

LANG

Legend:
- Word rate
- Sound power
- Pause duration
- End of sentence
- Frequency
- +J
- 5-gram surprisal
- PCFG surprisal
- Adaptive surprisal
- DLT

Axes: BOLD (y-axis), Time (x-axis)

LANG

MD

Legend:
- Word rate
- Sound power
- Pause duration
- End of sentence
- Frequency
- +J
- 5-gram surprisal
- PCFG surprisal
- Adaptive surprisal
- DLT
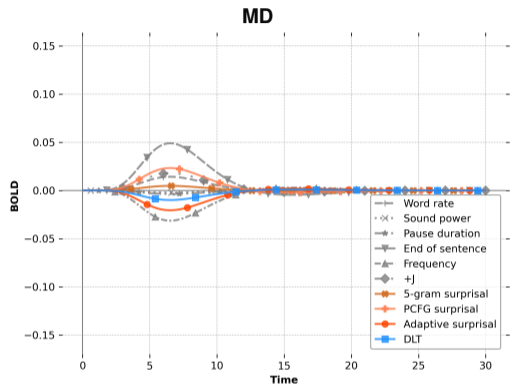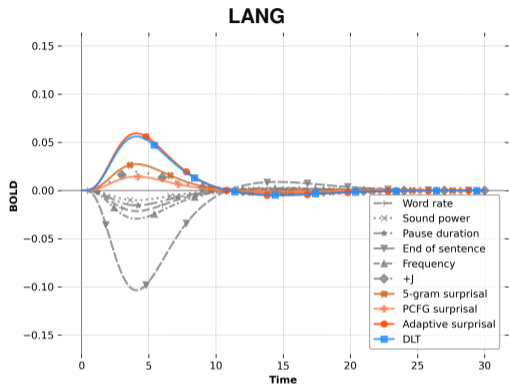
Large DLT effect in LANG, null/neg in MD
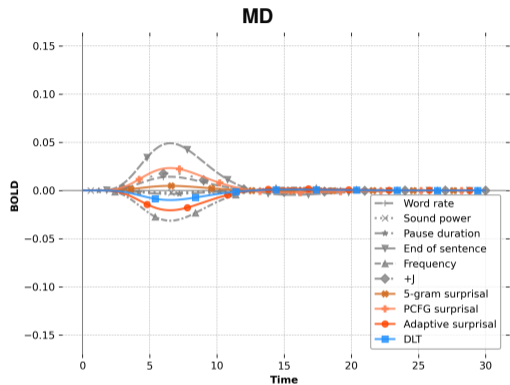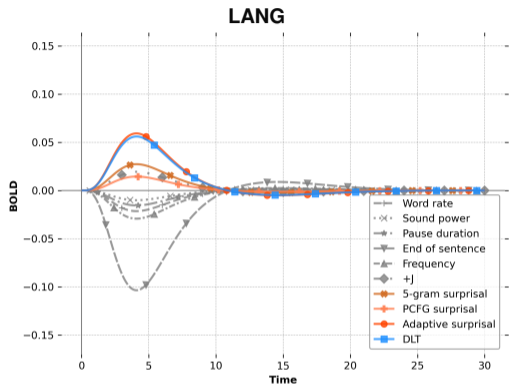
Large DLT effect in LANG, null/neg in MD

DLT significantly improves generalization in LANG but not MD
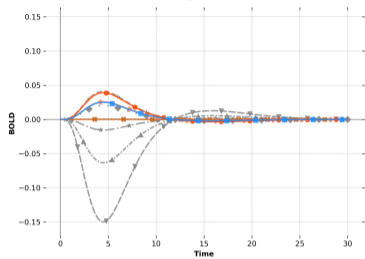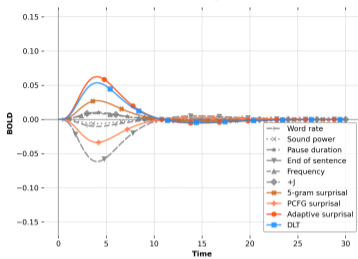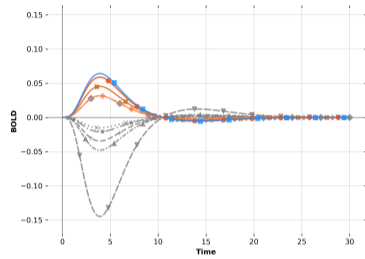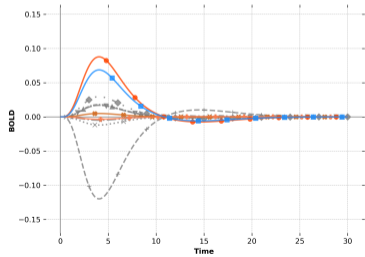
- - -	Word rate
⋯⋯	Sound power
- · -	Pause duration
—▼—	End of sentence
—▲—	Frequency
—◆—	+J
—●—	5-gram surprisal
—●—	PCFG surprisal
—●—	Adaptive surprisal
—■—	DLT
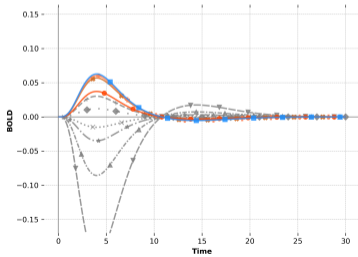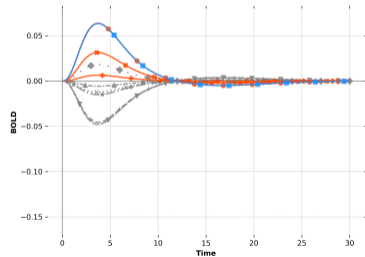
Q3: Does language comprehension involve memory retrieval?
Yes

Q3: Does language comprehension involve memory retrieval?
Yes

Q4: Are memory stores are domain-general?
No

+ Comprehenders represent and use syntax by default
+ Retrieval effects are not explained by prediction
+ Language processing is:

+ Comprehenders represent and use syntax by default
+ Retrieval effects are not explained by prediction
+ Language processing is:
    - Mostly carried out from domain general regions
    - Rarely uses language-specialised regions

+ Comprehenders represent and use syntax by default
+ Retrieval effects are not explained by prediction
+ Language processing is:
    + Mostly siloed off from domain-general regions
    + Distributed across language-specialized regions

+ Comprehenders represent and use syntax by default
+ Retrieval effects are not explained by prediction
+ Language processing is:
    + Mostly siloed off from domain-general regions
    + Distributed across language-specialized regions

+ Comprehenders represent and use syntax by default
+ Retrieval effects are not explained by prediction
+ Language processing is:
    + Mostly siloed off from domain-general regions
    + Distributed across language-specialized regions

Epilogue:
Naturalistic Language Processing in Reading Times vs. Imaging
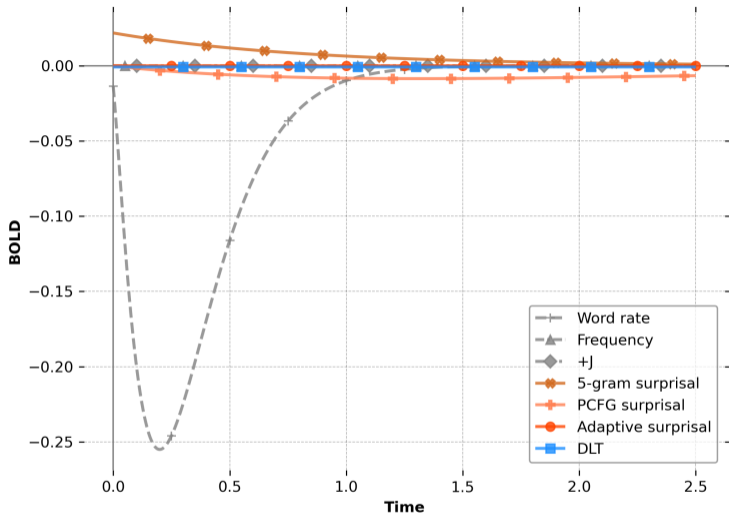
Self-paced reading

(Futrell et al. 2018)

Puzzle:

Puzzle:

Strong syntactic effects when reading constructed stimuli

(e.g. Grodner and Gibson 2005)

Puzzle:

Strong syntactic effects when reading constructed stimuli
(e.g. Grodner and Gibson 2005)

Weak/null syntactic effects when reading naturalistic stimuli
(e.g. Demberg and Keller 2008; Frank and Bod 2011; Schijndel and Schuler 2013)

Puzzle:

Strong syntactic effects when reading constructed stimuli
(e.g. Grodner and Gibson 2005)

Weak/null syntactic effects when reading naturalistic stimuli
(e.g. Demberg and Keller 2008; Frank and Bod 2011; Schijndel and Schuler 2013)

Strong syntactic effects **in brains** when listening to naturalistic stimuli
this study

Puzzle:

Strong syntactic effects when reading constructed stimuli
(e.g. Grodner and Gibson 2005)

Weak/null syntactic effects when reading naturalistic stimuli
(e.g. Demberg and Keller 2008; Frank and Bod 2011; Schijndel and Schuler 2013)

Strong syntactic effects **in brains** when listening to naturalistic stimuli
this study

Possible implication:
Some computational demands may not cause readers to slow down

References

Brennan, Jonathan et al. (2016). "Abstract linguistic structure correlates with temporal activity during naturalistic comprehension". In: Brain and Language 157, pp. 81–94.

Campbell, Karen L and Lorraine K Tyler (2018). "Language-related domain-specific and domain-general systems in the human brain". In: Current opinion in behavioral sciences 21, pp. 132–137.

Caplan, David and Gloria S Waters (1999). "Verbal working memory and sentence comprehension". In: Behavioral and Brain Sciences 22.1, pp. 77–94.

Demberg, Vera and Frank Keller (2008). "Data from eye-tracking corpora as evidence for theories of syntactic processing complexity". In: Cognition 109.2, pp. 193–210.

Diachek, Evgeniia et al. (2019). "The domain-general multiple demand (MD) network does not support core aspects of language comprehension: A large-scale fMRI investigation". In:

Duncan, John (2010). "The multiple-demand (MD) system of the primate brain: Mental programs for intelligent behaviour". In: Trends in Cognitive Sciences 14.4, pp. 172–179.

References

Federmeier, Kara D, Suzanne R Jongman, and Jakub M Szewczyk (2020). "Examining the Role of General Cognitive Skills in Language Processing: A Window Into Complex Cognition". In: Current Directions in Psychological Science, p. 0963721420964095.

Federmeier, Kara D et al. (2002). "The impact of semantic memory organization and sentence context information on spoken language processing by younger and older adults: An ERP study". In: Psychophysiology 39.2, pp. 133–146.

Fedorenko, Evelina and Idan A Blank (2020). "Broca's Area Is Not a Natural Kind". In: Trends in Cognitive Sciences.

Fedorenko, Evelina, Edward Gibson, and Douglas Rohde (2006). "The nature of working memory capacity in sentence comprehension: Evidence against domain-specific working memory resources". In: Journal of memory and language 54.4, pp. 541–553.

Fedorenko, Evelina et al. (2010). "New method for fMRI investigations of language: defining ROIs functionally in individual subjects". In: Journal of neurophysiology 104.2, pp. 1177–1194.

References

Ferreira, Fernanda and Suphasiree Chantavarin (2018). "Integration and prediction in language processing: A synthesis of old and new". In: Current directions in psychological science 27.6, pp. 443–448.

Fiebach, Christian J, Matthias Schlesewsky, and Angela D Friederici (2001). "Syntactic working memory and the establishment of filler-gap dependencies: Insights from ERPs and fMRI". In: Journal of psycholinguistic research 30.3, pp. 321–338.

Frank, Stefan L and Rens Bod (June 2011). "Insensitivity of the Human Sentence-Processing System to Hierarchical Structure". In: Psychological Science 22.6, pp. 829–834. ISSN: 0956-7976. DOI: 10.1177/0956797611409589. URL: http://www.ncbi.nlm.nih.gov/pubmed/21586764http://journals.sagepub.com/doi/10.1177/0956797611409589.

Friedrich, Roland and Angela D Friederici (2009). "Mathematical logic in the human brain: syntax". In: PloS one 4.5.

References

Frost, Martin A and Rainer Goebel (2012). "Measuring structural–functional correspondence: spatial variability of specialised brain regions after macro-anatomical alignment". In: Neuroimage 59.2, pp. 1369–1381.

Futrell, Richard et al. (May 2018). "The Natural Stories Corpus". In: Proceedings of the Eleventh International Conference on Language Resources and Evaluation. Ed. by Nicoletta Calzolari et al. Paris, France: European Language Resources Association (ELRA). ISBN: 979-10-95546-00-9.

Gambi, Chiara et al. (2018). "The development of linguistic prediction: Predictions of sound and meaning in 2- to 5-year-olds". In: Journal of Experimental Child Psychology 173, pp. 351–370. ISSN: 0022-0965. DOI: https://doi.org/10.1016/j.jecp.2018.04.012. URL: http://www.sciencedirect.com/science/article/pii/S0022096517307531.

Gibson, Edward (2000). "The dependency locality theory: A distance-based theory of linguistic complexity". In: Image, language, brain: Papers from the first mind articulation project symposium. Cambridge, MA: MIT {P}ress, pp. 95–126.

References

Grodner, Daniel J and Edward Gibson (2005). "Consequences of the serial nature of linguistic input". In: Cognitive Science 29, pp. 261–291.

Hale, John (2001). "A probabilistic Earley parser as a psycholinguistic model". In: Proceedings of the Second Meeting of the North American Chapter of the Association for Com pp. 1–8. DOI: 10.3115/1073336.1073357. URL: http://www.aclweb.org/anthology/N01-1021http://portal.acm.org/citation.cfm?doid=1073336.1073357.

Handwerker, Daniel A, John M Ollinger, and Mark D'Esposito (2004). "Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses.". In: NeuroImage 21.4, pp. 1639–1651. URL: http://www.ncbi.nlm.nih.gov/pubmed/15050587.

Hasson, Uri and Christopher J Honey (2012). "Future trends in Neuroimaging: Neural processes as expressed within real-life contexts". In: NeuroImage 62.2, pp. 1272–1278.

Hasson, Uri et al. (2018). "Grounding the neurobiology of language in first principles: The necessity of non-language-centric explanations for language comprehension". In: Cognition 180, pp. 135–157.

References

Henderson, John M et al. (2016). "Language structure in the brain: A fixation-related fMRI study of syntactic surprisal in reading". In: Neuroimage 132, pp. 293–300.

Huettig, Falk and Nivedita Mani (2016). "Is prediction necessary to understand language? Probably not". In: Language, Cognition and Neuroscience 31.1, pp. 19–31.

January, David, John C Trueswell, and Sharon L Thompson-Schill (2009). "Co-localization of Stroop and syntactic ambiguity resolution in Broca's area: Implications for the neural basis of sentence processing". In: Journal of Cognitive Neuroscience 21.12, pp. 2434–2444.

Kuperberg, Gina R and T Florian Jaeger (2016). "What do we mean by prediction in language comprehension?" In: Language, cognition and neuroscience 31.1, pp. 32–59.

Kuperberg, Gina R et al. (2003). "Distinct patterns of neural modulation during the processing of conceptual and syntactic anomalies". In: Journal of Cognitive Neuroscience 15.2, pp. 272–293.

Levy, Roger (2008). "Expectation-based syntactic comprehension". In: Cognition 106.3, pp. 1126–1177.

References

Lewis, Richard L and Shravan Vasishth (2005). "An activation-based model of sentence processing as skilled memory retrieval". In: Cognitive Science 29.3, pp. 375–419.

Linck, Jared A et al. (2014). "Working memory and second language comprehension and production: A meta-analysis". In: Psychonomic bulletin \& review 21.4, pp. 861–883.

Lopopolo, Alessandro et al. (2017). "Using stochastic language models (SLM) to map lexical, syntactic, and phonological information processing in the brain". In: PloS one 12.5, e0177794.

Lopopolo, Alessandro et al. (2020). "Distinguishing syntactic operations in the brain: Dependency and phrase-structure parsing". In: Neurobiology of Language Just Accepted, pp. 1–64.

Mani, Nivedita and Falk Huettig (2012). "Prediction during language processing is a piece of cake—But only for skilled producers.". In: Journal of Experimental Psychology: Human Perception and Performance 38.4, p. 843.

References

Novick, Jared M, John C Trueswell, and Sharon L Thompson-Schill (2005). "Cognitive control and parsing: Reexamining the role of Broca's area in sentence comprehension". In: Cognitive, Affective, \& Behavioral Neuroscience 5.3, pp. 263–281.

Patel, Aniruddh D (2003). "Language, music, syntax and the brain". In: Nature Neuroscience 6.7, pp. 674–681.

Pickering, Martin J and Chiara Gambi (2018). "Predicting while comprehending language: A theory and review.". In: Psychological bulletin 144.10, p. 1002.

Poldrack, Russell A (2006). "Can cognitive processes be inferred from neuroimaging data?" In: Trends in cognitive sciences 10.2, pp. 59–63.

Rasmussen, Nathan E and William Schuler (2018). "Left-Corner Parsing With Distributed Associative Memory Produces Surprisal and Locality Effects". In: Cognitive Science 42, pp. 1009–1042.

Schijndel, Marten van and William Schuler (2013). "An Analysis of Frequency- and Memory-Based Processing Costs". In: Proceedings of Human Language Technologies: The 2013 Annual Conference of the North Am

References

Shain, Cory and William Schuler (2018). "Deconvolutional time series regression: A technique for modeling temporally diffuse effects". In: Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pp. 2679–2689.

– (2019). "Continuous-Time Deconvolutional Regression for Psycholinguistic Modeling". In: PsyArXiv.

Smith, Nathaniel J and Roger Levy (2013). "The effect of word predictability on reading time is logarithmic". In: Cognition 128, pp. 302–319.

Stowe, Laurie A et al. (1998). "Localizing components of a complex task: Sentence processing and working memory". In: Neuroreport 9.13, pp. 2995–2999.

Strijkers, Kristof et al. (2019). "Grammatical class modulates the (left) inferior frontal gyrus within 100 milliseconds when syntactic context is predictive". In: Scientific reports 9.1, p. 4830.

Swets, Benjamin et al. (2008). "Underspecification of syntactic ambiguities: Evidence from self-paced reading". In: Memory and Cognition 36.1, pp. 201–216.

References

Tettamanti, Marco and Dorothea Weniger (2006). "Broca's area: A supramodal hierarchical processor?" In: Cortex 42.4, pp. 491–494.

van Schijndel, Marten and Tal Linzen (2018). "A Neural Model of Adaptation in Reading". In: EMNLP 2018, pp. 4704–4710.

van Schijndel, Marten and William Schuler (2013). "An Analysis of Frequency- and Memory-Based Processing Costs". In: Proceedings of Human Language Technologies: The 2013 Annual Conference of the North American Association for Computational Linguistics.

Willems, Roel M et al. (2015). "Prediction during natural language comprehension". In: Cerebral Cortex 26.6, pp. 2506–2516.

+ Linguistic prediction may be carried out by domain-general executive control
  (Linck et al. 2014; Huettig and Mani 2016; Pickering and Gambi 2018; Strijkers et al. 2019)

  + Hierarchic sequential prediction is a domain-general skill
    (**vanschijndeletal13:tcs2**; Smith and Levy 2013; Rasmussen and Schuler 2018)

  + Variation in executive function modulates prediction effects
    (Federmeier et al. 2002; Mani and Huettig 2012; Gambi et al. 2018)

  + Domain-general executive regions engage during some language tasks
    (Kuperberg et al. 2003; Novick et al. 2005; January et al. 2009, cf. Diachek et al. 2019)

  + Broca's may be universal syntactic processor
    (Patel 2003; Tettamanti and Weniger 2006; Friedrich and Friederici 2009, cf. Fedorenko and Blank 2020)

+ Plausible executive resource: fronto-parietal multiple demand (MD) network
  (Duncan 2010)

+ **Hypothesis 2:** Linguistic prediction recruits MD

+ Linguistic prediction may be carried out by domain-general executive control
  (Linck et al. 2014; Huettig and Mani 2016; Pickering and Gambi 2018; Strijkers et al. 2019)

    + Hierarchic sequential prediction is a domain-general skill
      (**vanschijndeletal13:tcs2**; Smith and Levy 2013; Rasmussen and Schuler 2018)

    + Variation in executive function modulates prediction effects
      (Federmeier et al. 2002; Mani and Huettig 2012; Gambi et al. 2018)

    + Domain-general executive regions engage during some language tasks
      (Kuperberg et al. 2003; Novick et al. 2005; January et al. 2009, cf. Diachek et al. 2019)

    + Broca's may be universal syntactic processor
      (Patel 2003; Tettamanti and Weniger 2006; Friedrich and Friederici 2009, cf. Fedorenko and Blank 2020)

+ Plausible executive resource: fronto-parietal multiple demand (MD) network
  (Duncan 2010)

+ **Hypothesis 2:** Linguistic prediction recruits MD

- Linguistic prediction may be carried out by domain-general executive control
  (Linck et al. 2014; Huettig and Mani 2016; Pickering and Gambi 2018; Strijkers et al. 2019)
    - Hierarchic sequential prediction is a domain-general skill
      (**vanschijndeletal13:tcs2**; Smith and Levy 2013; Rasmussen and Schuler 2018)
    - Variation in executive function modulates prediction effects
      (Federmeier et al. 2002; Mani and Huettig 2012; Gambi et al. 2018)
    - Domain-general executive regions engage during some language tasks
      (Kuperberg et al. 2003; Novick et al. 2005; January et al. 2009, cf. Diachek et al. 2019)
    - Broca's may be universal syntactic processor
      (Patel 2003; Tettamanti and Weniger 2006; Friedrich and Friederici 2009, cf. Fedorenko and Blank 2020)
- Plausible executive resource: fronto-parietal multiple demand (MD) network
  (Duncan 2010)
- **Hypothesis 2:** Linguistic prediction recruits MD

+ Linguistic prediction may be carried out by domain-general executive control
  (Linck et al. 2014; Huettig and Mani 2016; Pickering and Gambi 2018; Strijkers et al. 2019)
    + Hierarchic sequential prediction is a domain-general skill
      (**vanschijndeletal13:tcs2**; Smith and Levy 2013; Rasmussen and Schuler 2018)
    + Variation in executive function modulates prediction effects
      (Federmeier et al. 2002; Mani and Huettig 2012; Gambi et al. 2018)
    + Domain-general executive regions engage during some language tasks
      (Kuperberg et al. 2003; Novick et al. 2005; January et al. 2009, cf. Diachek et al. 2019)
    + Broca's may be universal syntactic processor
      (Patel 2003; Tettamanti and Weniger 2006; Friedrich and Friederici 2009, cf. Fedorenko and Blank 2020)

+ Plausible executive resource: fronto-parietal multiple demand (MD) network
  (Duncan 2010)

+ **Hypothesis 2:** Linguistic prediction recruits MD

- Linguistic prediction may be carried out by domain-general executive control
  (Linck et al. 2014; Huettig and Mani 2016; Pickering and Gambi 2018; Strijkers et al. 2019)
    - Hierarchic sequential prediction is a domain-general skill
      (**vanschijndeletal13:tcs2**; Smith and Levy 2013; Rasmussen and Schuler 2018)
    - Variation in executive function modulates prediction effects
      (Federmeier et al. 2002; Mani and Huettig 2012; Gambi et al. 2018)
    - Domain-general executive regions engage during some language tasks
      (Kuperberg et al. 2003; Novick et al. 2005; January et al. 2009, cf. Diachek et al. 2019)
    - Broca's may be universal syntactic processor
      (Patel 2003; Tettamanti and Weniger 2006; Friedrich and Friederici 2009, cf. Fedorenko and Blank 2020)
- Plausible executive resource: fronto-parietal multiple demand (MD) network
  (Duncan 2010)
- **Hypothesis 2:** Linguistic prediction recruits MD

+ Linguistic prediction may be carried out by domain-general executive control
  (Linck et al. 2014; Huettig and Mani 2016; Pickering and Gambi 2018; Strijkers et al. 2019)
    + Hierarchic sequential prediction is a domain-general skill
      (**vanschijndeletal13:tcs2**; Smith and Levy 2013; Rasmussen and Schuler 2018)
    + Variation in executive function modulates prediction effects
      (Federmeier et al. 2002; Mani and Huettig 2012; Gambi et al. 2018)
    + Domain-general executive regions engage during some language tasks
      (Kuperberg et al. 2003; Novick et al. 2005; January et al. 2009, cf. Diachek et al. 2019)
    + Broca's may be universal syntactic processor
      (Patel 2003; Tettamanti and Weniger 2006; Friedrich and Friederici 2009, cf. Fedorenko and Blank 2020)
+ Plausible executive resource: fronto-parietal multiple demand (MD) network
  (Duncan 2010)
+ **Hypothesis 2:** Linguistic prediction recruits MD

+ Linguistic prediction may be carried out by domain-general executive control
  (Linck et al. 2014; Huettig and Mani 2016; Pickering and Gambi 2018; Strijkers et al. 2019)
    + Hierarchic sequential prediction is a domain-general skill
      (**vanschijndeletal13:tcs2**; Smith and Levy 2013; Rasmussen and Schuler 2018)
    + Variation in executive function modulates prediction effects
      (Federmeier et al. 2002; Mani and Huettig 2012; Gambi et al. 2018)
    + Domain-general executive regions engage during some language tasks
      (Kuperberg et al. 2003; Novick et al. 2005; January et al. 2009, cf. Diachek et al. 2019)
    + Broca's may be universal syntactic processor
      (Patel 2003; Tettamanti and Weniger 2006; Friedrich and Friederici 2009, cf. Fedorenko and Blank 2020)
+ Plausible executive resource: fronto-parietal multiple demand (MD) network
  (Duncan 2010)
+ **Hypothesis 2:** Linguistic prediction recruits MD

OOS hypothesis tests: COMBINED

Baseline model
(both interactions ablated)

OOS hypothesis tests: COMBINED

Baseline model
(both interactions ablated)

p = 0.0001***

5-gram:Network model
(PCFG:Network ablated)

OOS hypothesis tests: COMBINED

PCFG:Network model
(5-gram:Network ablated)

p = 0.0001***
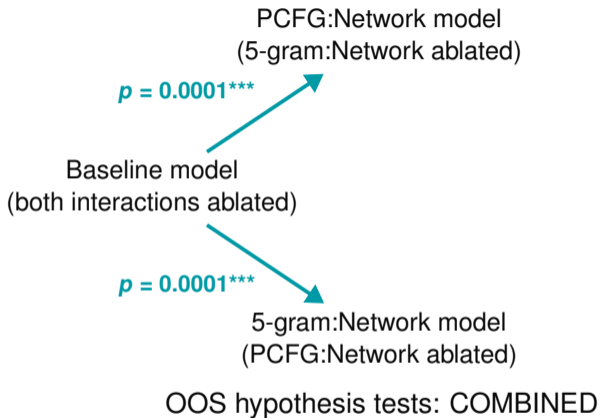
Baseline model
(both interactions ablated)

p = 0.0001***

5-gram:Network model
(PCFG:Network ablated)

OOS hypothesis tests: COMBINED

PCFG:Network model
(5-gram:Network ablated)

p = 0.0001***

Baseline model
(both interactions ablated)

Both model
(both interactions included)

p = 0.0001***

p = 0.0001***

5-gram:Network model
(PCFG:Network ablated)

OOS hypothesis tests: COMBINED

PCFG:Network model
(5-gram:Network ablated)

p = 0.0001***   *p* = 0.0001***

Baseline model
(both interactions ablated)

Both model
(both interactions included)

p = 0.0001***   *p* = 0.0001***

5-gram:Network model
(PCFG:Network ablated)

OOS hypothesis tests: COMBINED

PCFG:Network model
(5-gram:Network ablated)

p = 0.0001***   *p* = 0.0001***

Baseline model   Both model
(both interactions ablated)   (both interactions included)

p = 0.0001***   *p* = 0.0001***

5-gram:Network model
(PCFG:Network ablated)

OOS hypothesis tests: COMBINED
Both 5-gram and PCFG surprisal effects are larger in LANG than MD.