Acquiring *Wh*-Dependencies Through Efficient Representation

Niels Dickson

UCI Language Science Community Talk
May 23, 2023

Acquiring *Wh*-Dependencies

Efficient Representation

Acquiring *Wh*-Dependencies

Efficient Representation

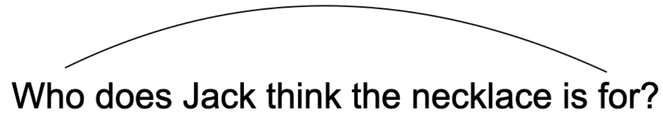
Acquiring *Wh*-Dependencies

Who does Jack think the necklace is for?

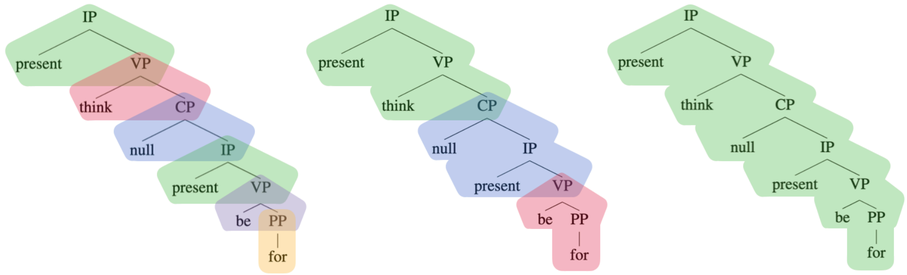
Efficient Representation

Acquiring

Wh-Dependencies



Efficient Representation

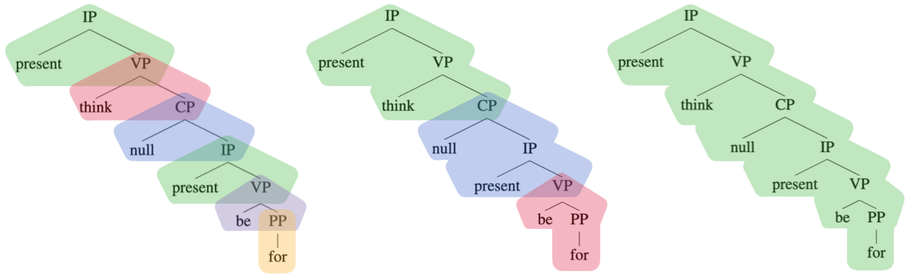


Acquiring *Wh*-Dependencies

Who does Jack think the necklace is for?

?

Efficient Representation



wh-dependencies

wh-dependencies

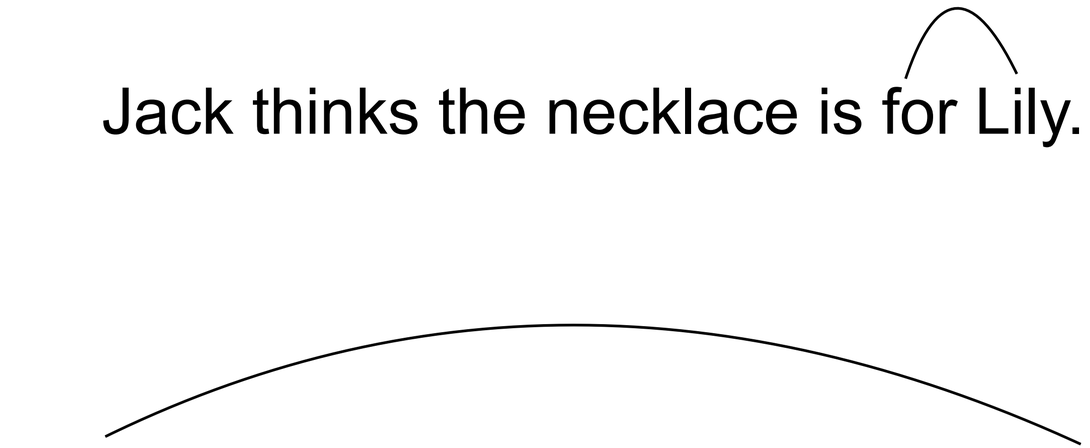
Jack thinks the necklace is for Lily.

wh-dependencies

wh-dependencies

Who does Jack think the necklace is for?

wh-dependencies



constraints on *wh*-dependencies

constraints on *wh*-dependencies

Jack thinks the necklace for Lily is expensive.

constraints on *wh*-dependencies

constraints on *wh*-dependencies

Jack thinks the necklace for Lily is expensive.

Who does Jack think [the necklace for] is expensive?

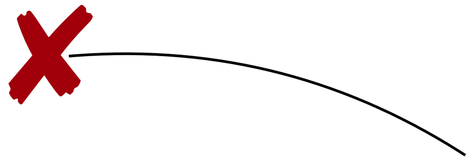
constraints on *wh*-dependencies

constraints on *wh*-dependencies

Jack thinks the necklace for Lily is expensive.

Who does Jack think [the necklace for] is expensive?

constraints on *wh*-dependencies



Subject Island: Who does Jack think [the necklace for] is expensive?

- **Complex NP Island:** What did Lily make the claim that Jack forgot?
- **Whether Island:** What does the teacher wonder whether Jack stole?
- **Adjunct Island:** What does the teach worry if Lily forgot?

roadmap

roadmap

- output: behavioral patterns for these dependencies

roadmap

- output: behavioral patterns for these dependencies
  - pattern 1: island effect
  - pattern 2: verb frequency effect
  - pattern 3: child preferences

roadmap

- output: behavioral patterns for these dependencies
  - pattern 1: island effect
  - pattern 2: verb frequency effect
  - pattern 3: child preferences
- input: what do children hear and what input are we feeding our model

roadmap

- output: behavioral patterns for these dependencies
  - pattern 1: island effect
  - pattern 2: verb frequency effect
  - pattern 3: child preferences
- input: what do children hear and what input are we feeding our model
- learning objective: identifying an efficient representation of the *wh*-dependency input

roadmap

- output: behavioral patterns for these dependencies
  - pattern 1: island effect
  - pattern 2: verb frequency effect
  - pattern 3: child preferences
- input: what do children hear and what input are we feeding our model
- learning objective: identifying an efficient representation of the *wh*-dependency input
- results: a learner adopting this representation strategy successfully attains the target knowledge

roadmap

- output: behavioral patterns for these dependencies
  - pattern 1: island effect
  - pattern 2: verb frequency effect
  - pattern 3: child preferences
- input: what do children hear and what input are we feeding our model
- learning objective: identifying an efficient representation of the *wh*-dependency input
- results: a learner adopting this representation strategy successfully attains the target knowledge
- discussion and future direction

pattern 1

island effect stimuli

	Main	Embedded
Non-Island		
Island		

Sprouse et al. 2012

pattern 1

island effect stimuli

	Main	Embedded
Non-Island	Who __ thinks the necklace is expensive?	
Island		

Sprouse et al. 2012

pattern 1

island effect stimuli

	Main	**Embedded**
Non-Island	Who __ thinks the necklace is expensive?	What does Jack think __ is expensive?
Island		

Sprouse et al. 2012

pattern 1

island effect stimuli

	Main	Embedded
Non-Island	Who __ thinks the necklace is expensive?	What does Jack think __ is expensive?
Island	Who __ thinks the necklace for Lily is expensive?	Who does Jack think the necklace for __ is expensive?

Sprouse et al. 2012

pattern 1

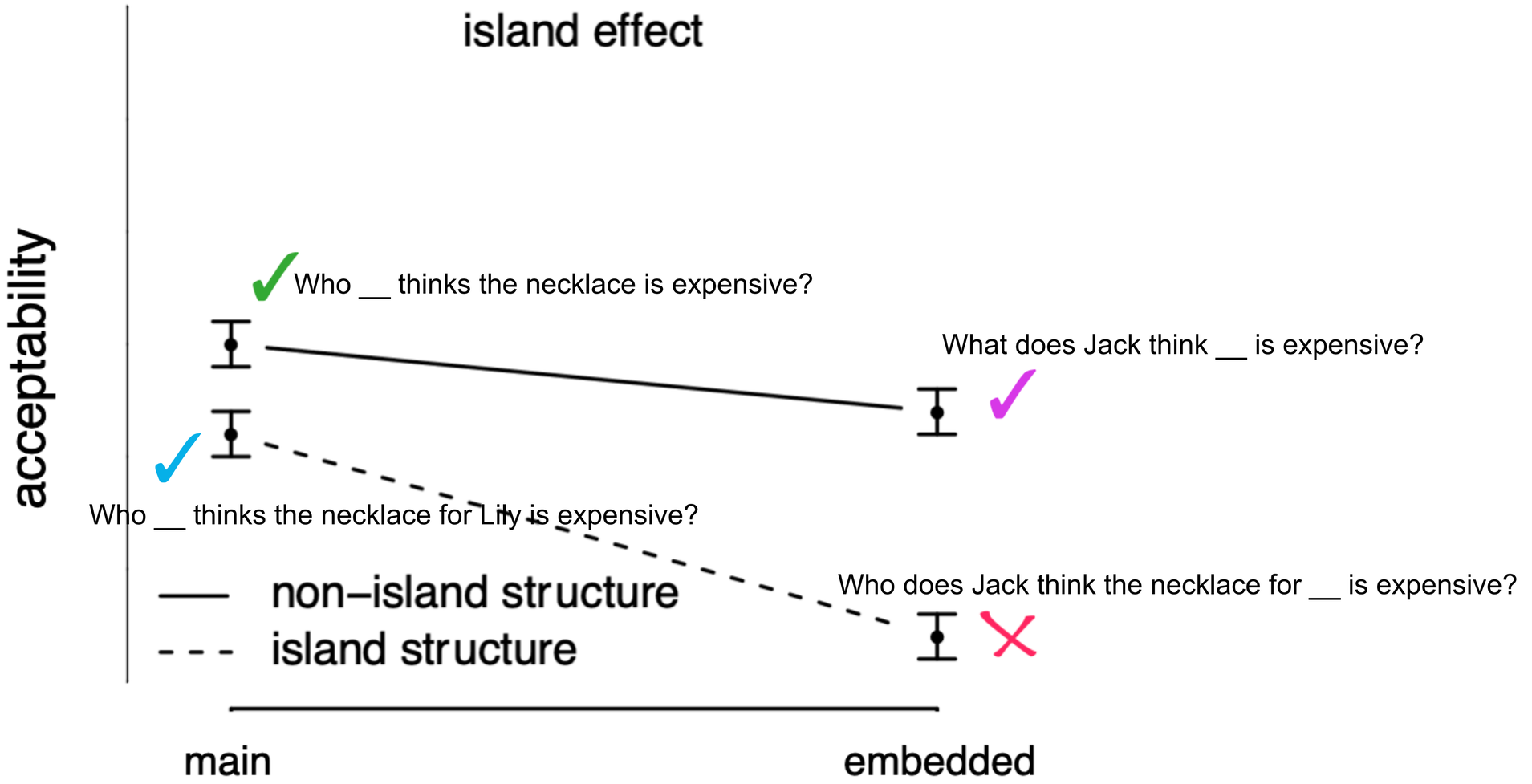
island effect stimuli

	Main	Embedded
Non-Island	✓ Who __ thinks the necklace is expensive?	✓ What does Jack think __ is expensive?
Island	✓ Who __ thinks the necklace for Lily is expensive?	✗ Who does Jack think the necklace for __ is expensive?

Sprouse et al. 2012

pattern 1

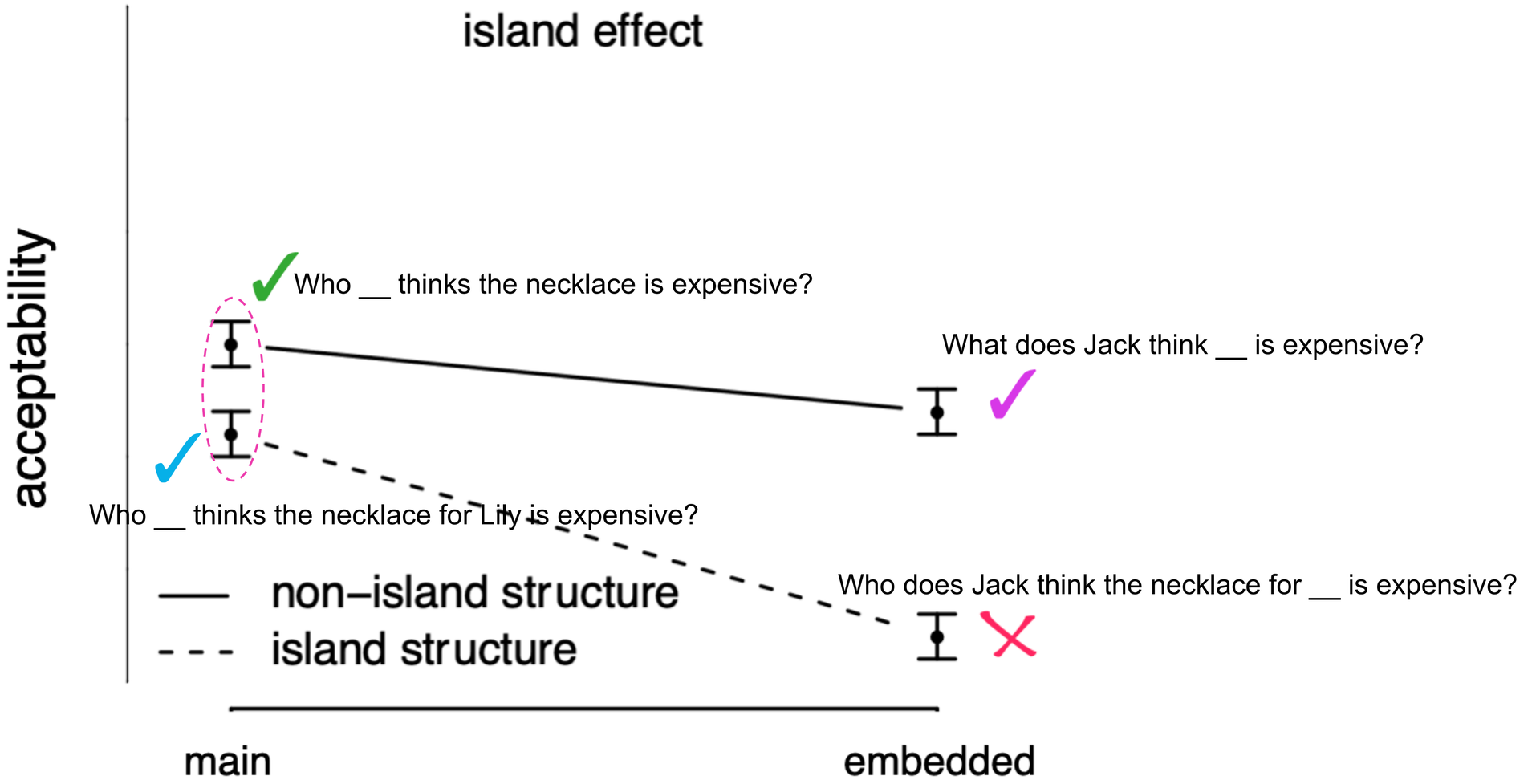
super-additive pattern



Sprouse et al. 2012

pattern 1

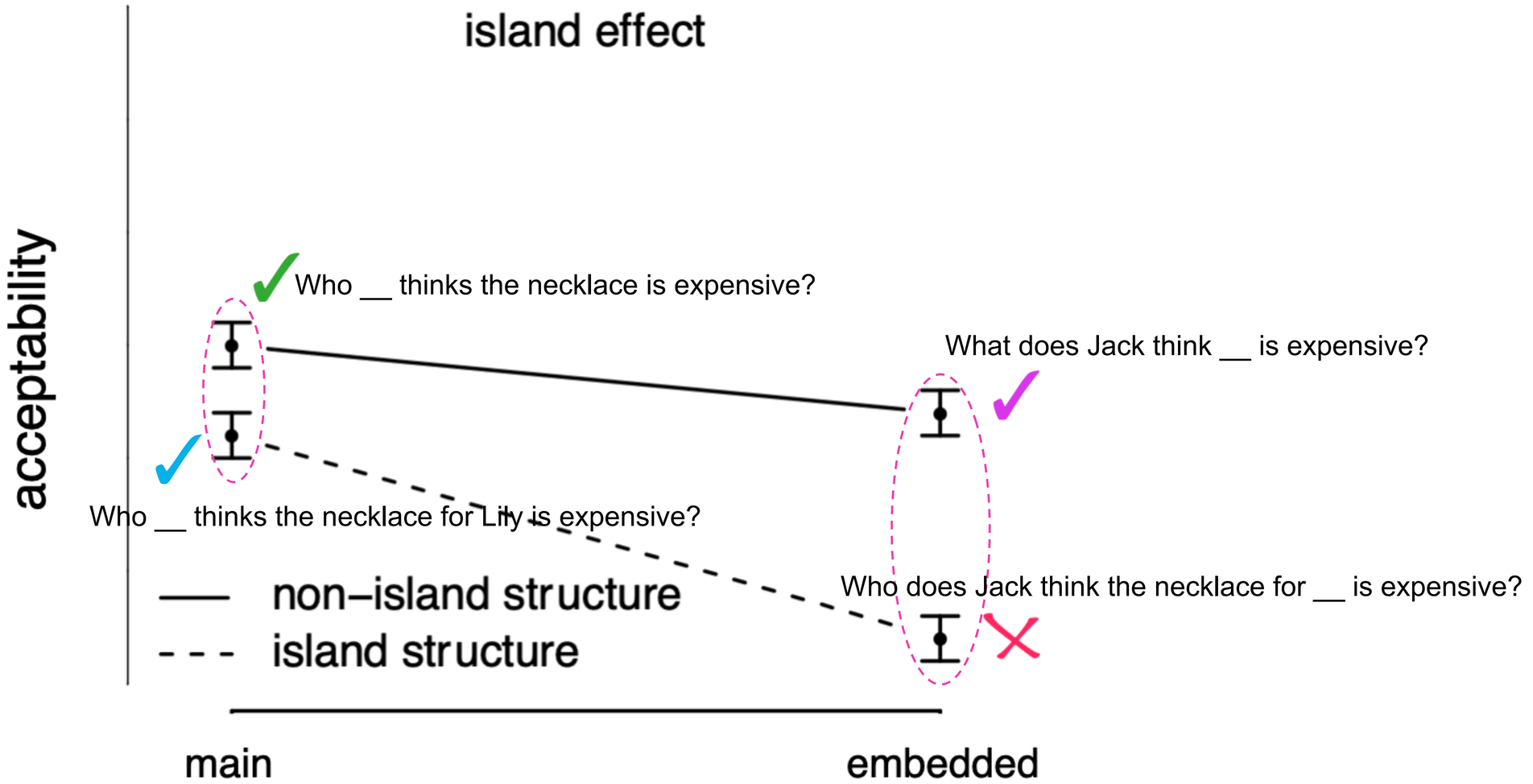
super-additive pattern



Sprouse et al. 2012

pattern 1

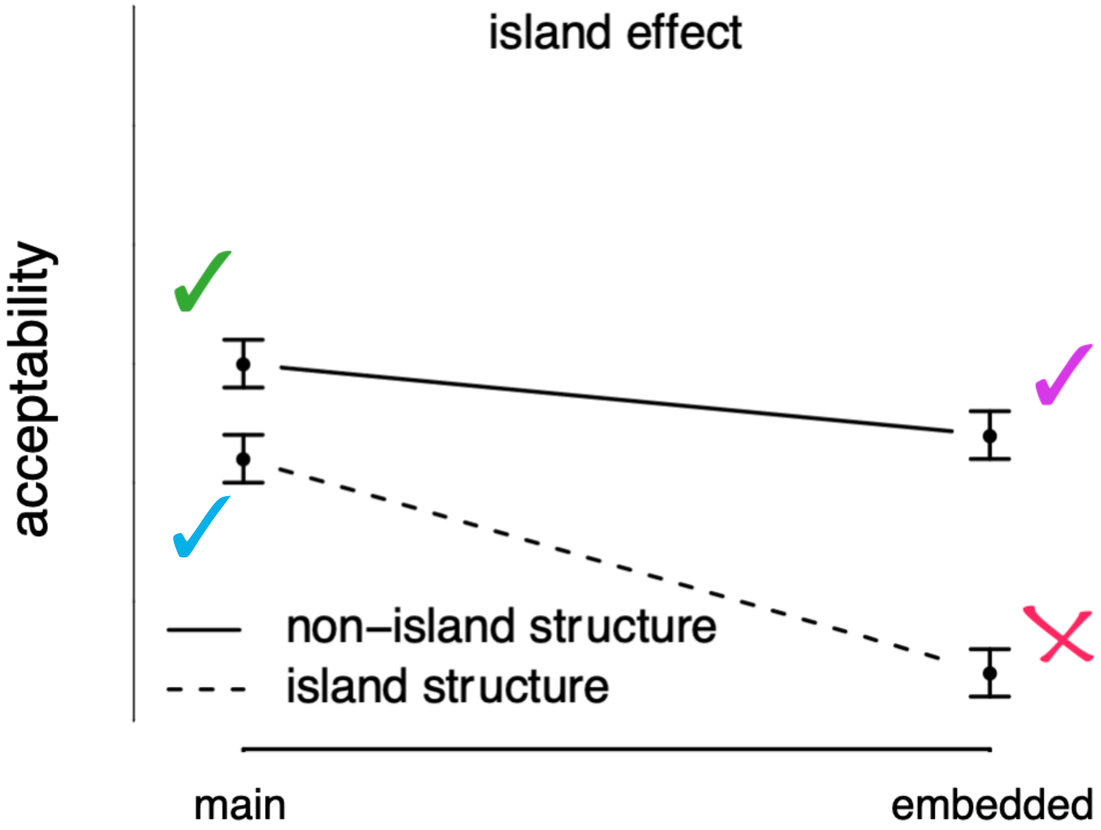
super-additive pattern

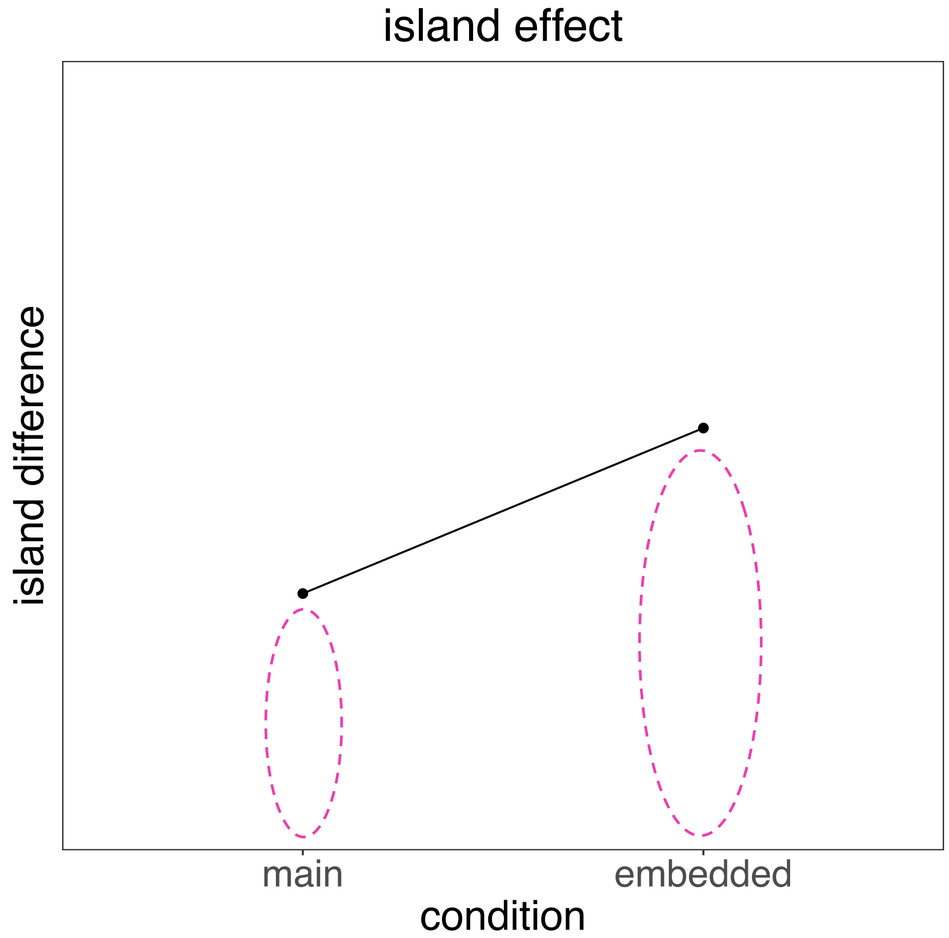


Sprouse et al. 2012

pattern 1

super-additive pattern

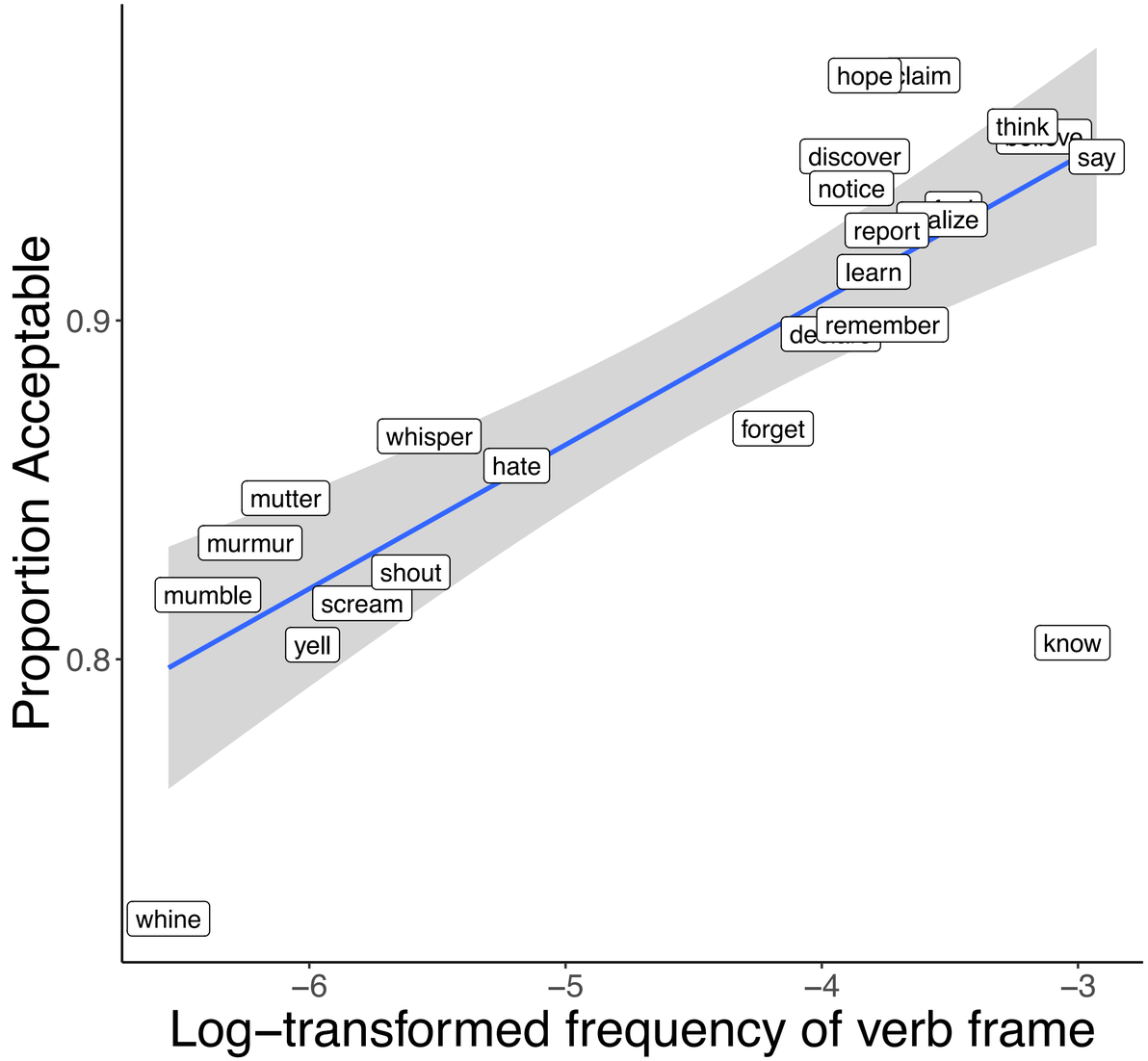




Sprouse et al. 2012

pattern 2

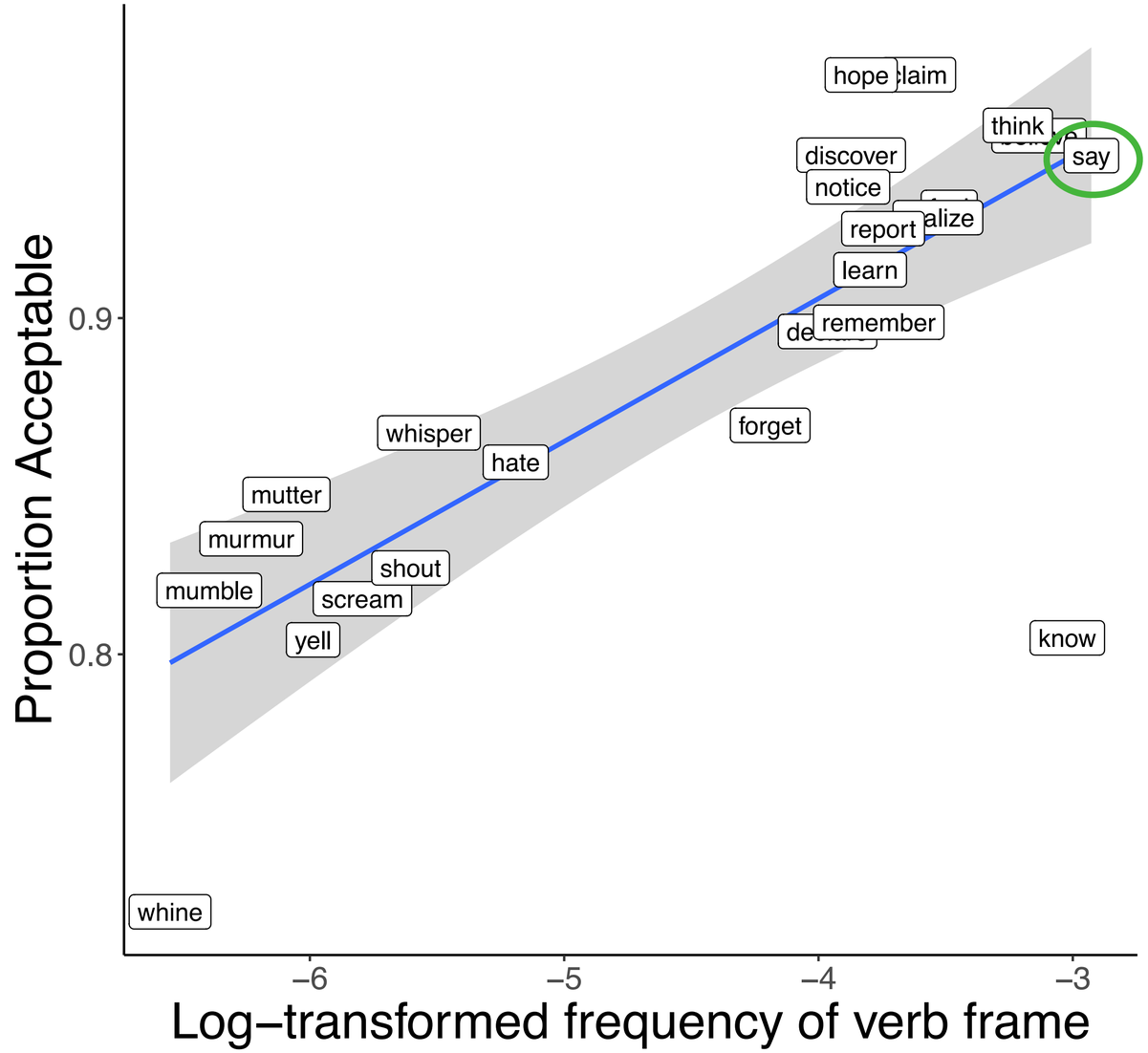
influence of verb frequency



Liu et al. 2022

pattern 2

influence of verb frequency

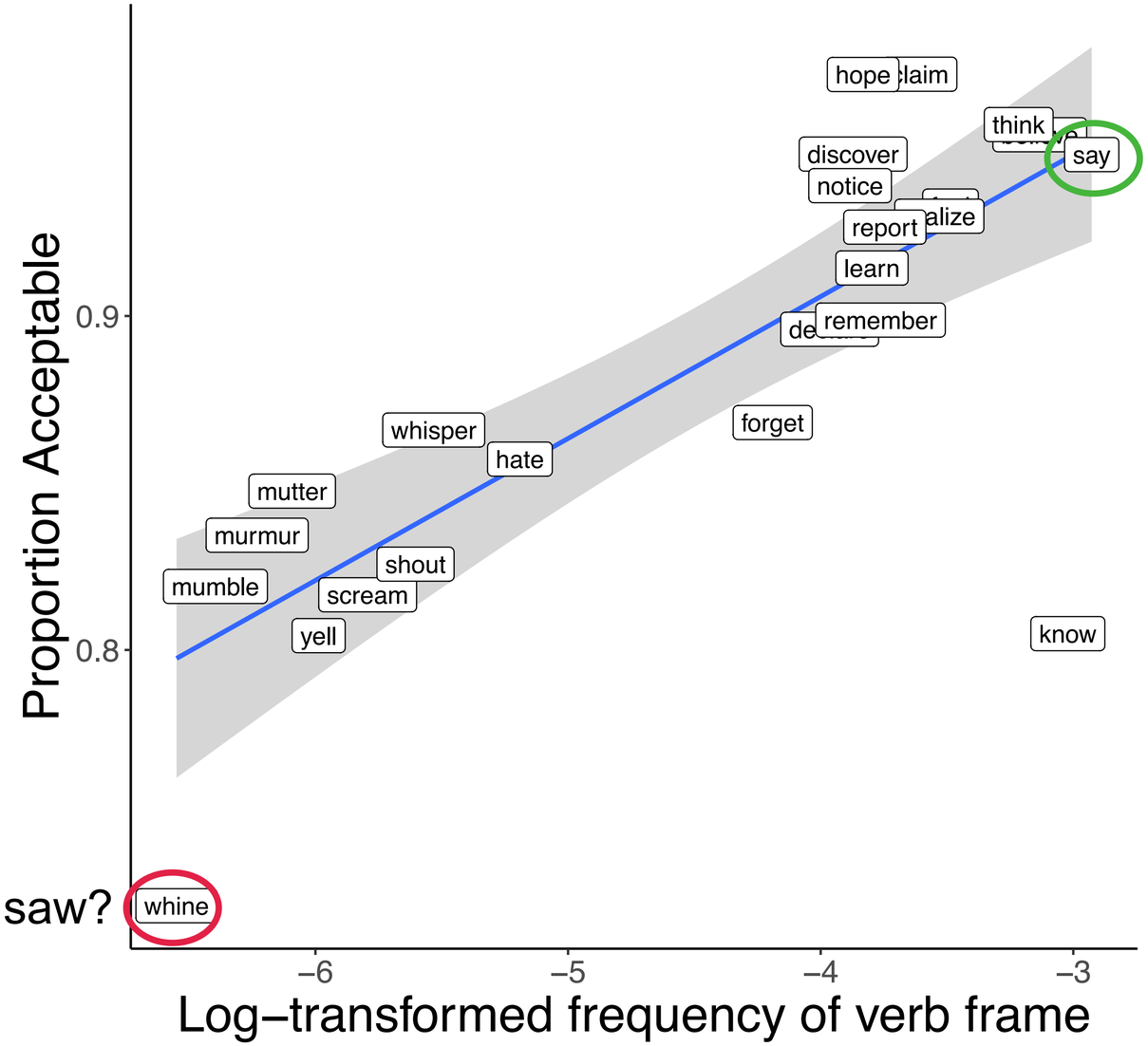


What did Lily say that he saw?

Liu et al. 2022

pattern 2

influence of verb frequency



What did Lily say that he saw?

What did Lily whine that he saw?

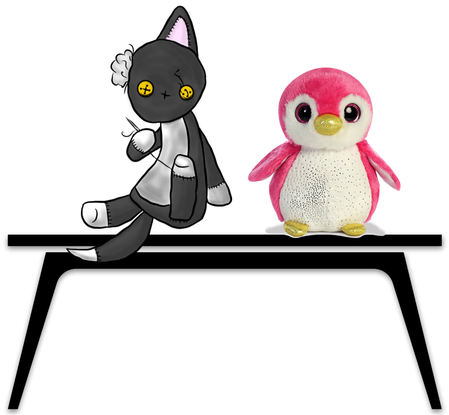
Liu et al. 2022

pattern 3

child interpretation preferences

How do children prefer to interpret potentially ambiguous *wh*-questions?

context



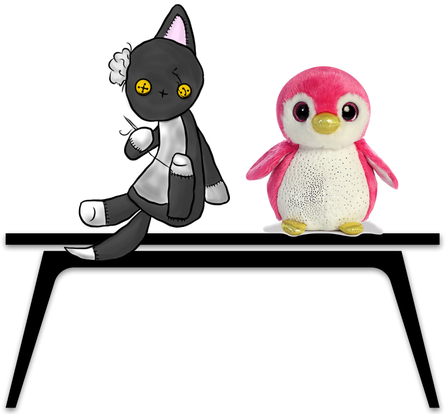
De Villiers et al. 2008

pattern 3

child interpretation preferences

How do children prefer to interpret potentially ambiguous *wh*-questions?

What did the boy fix the cat that was lying on the table with ___*what*?



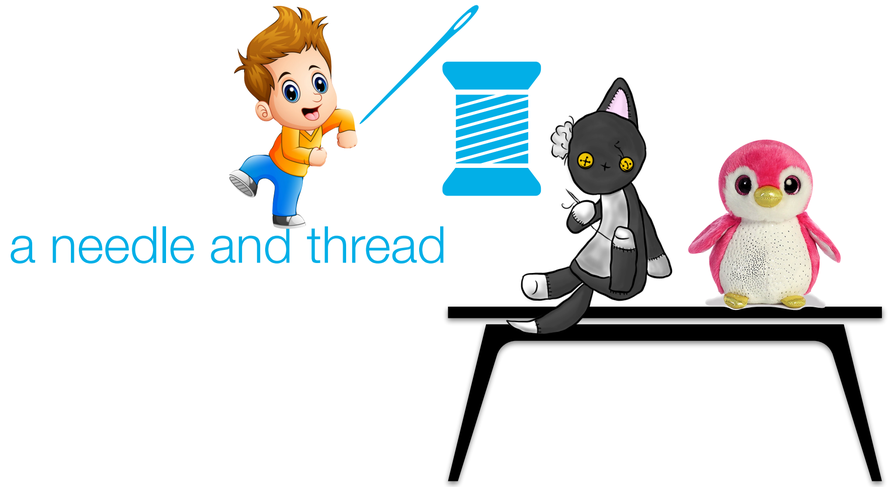
De Villiers et al. 2008

pattern 3

child interpretation preferences

How do children prefer to interpret potentially ambiguous *wh*-questions?

What did the boy [fix the cat *that was lying on the table* [with ___*what*]]?



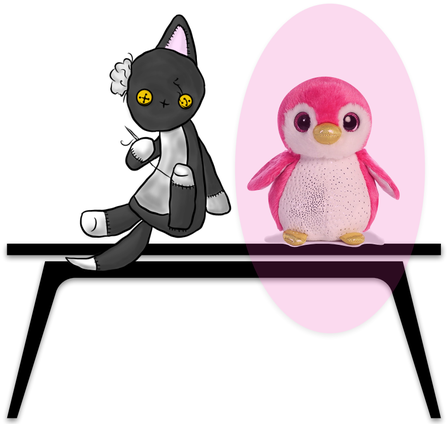
De Villiers et al. 2008

pattern 3

child interpretation preferences

How do children prefer to interpret potentially ambiguous *wh*-questions?

What did the boy [fix [the cat [that [was [lying [on [the table [with ___*what*]]]]]]]]?



a penguin

De Villiers et al. 2008

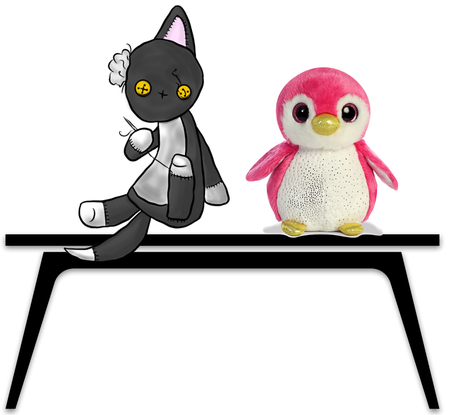
pattern 3

child interpretation preferences

How do children prefer to interpret potentially ambiguous *wh*-questions?

What did the boy fix the cat that was lying on the table with ___*what*?

children strongly prefer this interpretation



De Villiers et al. 2008

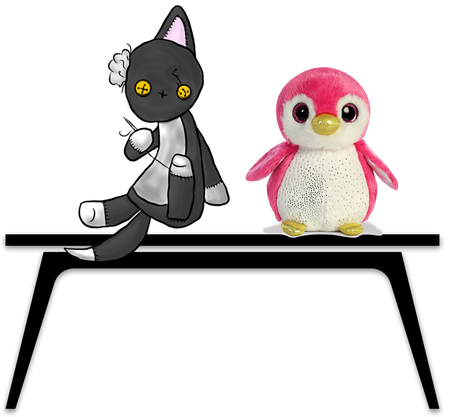
pattern 3

child interpretation preferences

How do children prefer to interpret potentially ambiguous *wh*-questions?

What did the boy [fix [the cat [that [was [lying [on [the table [with ___*what*]]]]]]]]?

This means they strongly disprefer the *wh*-dependency this interpretation relies on.



De Villiers et al. 2008

pattern 3

child interpretation preferences

Utterance	How often children preferred the longer *wh*-dependency
How did the boy say he hurt himself?	0.80
What did the mother say she bought?	0.79
Who did the police woman help to call?	0.48
Who did the little sister ask how to see?	0.25
How did the boy who sneezed drink the milk?	0.20
What did the boy fix the cat that was lying on the table with?	0.09
How did the girl ask where to ride?	0.04
Who did the boy ask what to bring?	0.04
How did the mom learn what to bake?	0.03

De Villiers et al. 2008

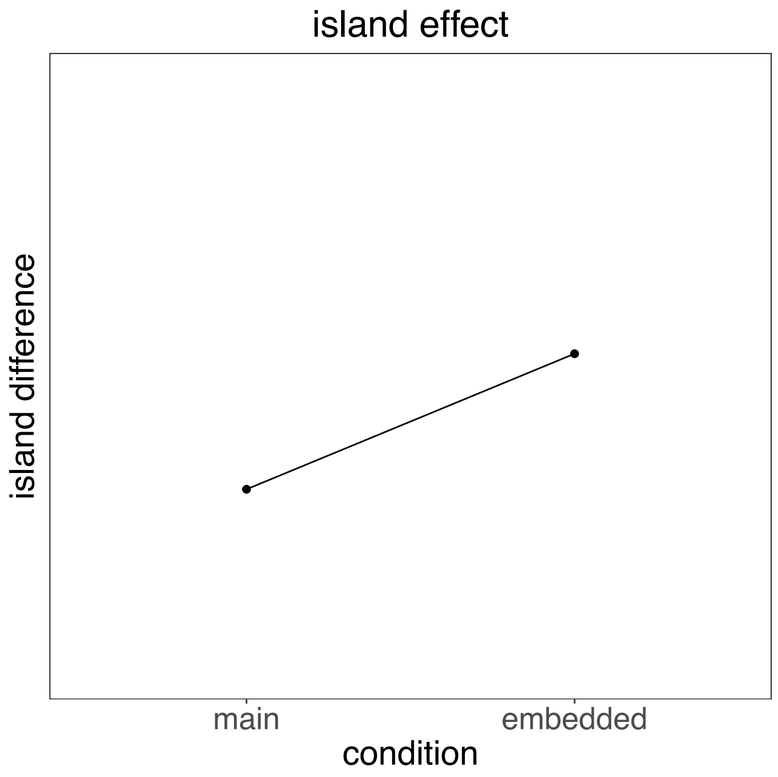
pattern 3

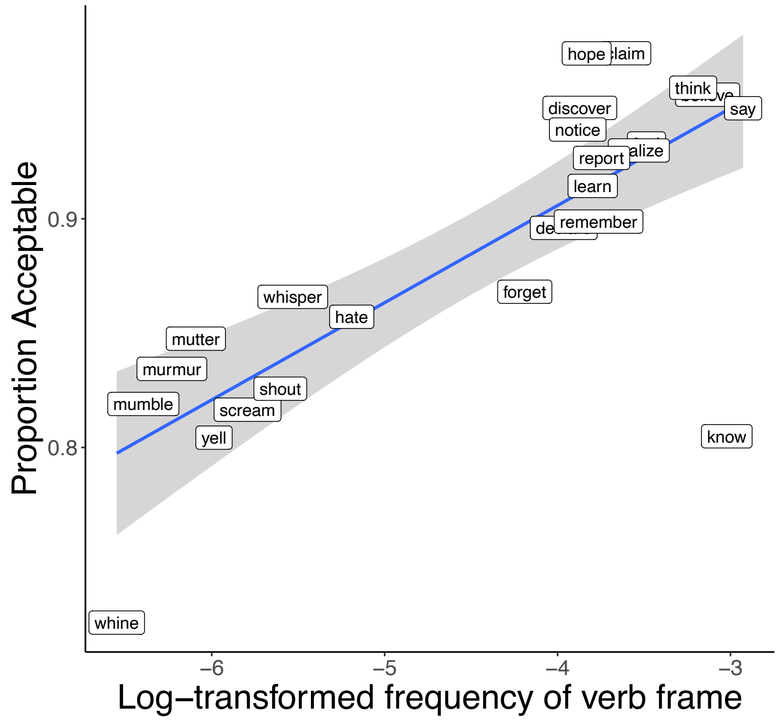
child interpretation preferences

Utterance	How often children preferred the longer *wh*-dependency	Longer dependency preferred
How did the boy say he hurt himself?	0.80	✓
What did the mother say she bought?	0.79	✓
Who did the police woman help to call?	0.48	✗
Who did the little sister ask how to see?	0.25	✗
How did the boy who sneezed drink the milk?	0.20	✗
What did the boy fix the cat that was lying on the table with?	0.09	✗
How did the girl ask where to ride?	0.04	✗
Who did the boy ask what to bring?	0.04	✗
How did the mom learn what to bake?	0.03	✗

De Villiers et al. 2008

behavioral patterns





Longer dependency preferred
✓
✓
✗
✗
✗
✗
✗
✗
✗

roadmap

- output: behavioral patterns for these dependencies
- input: what do children hear and what input are we feeding our model
- learning objective: identifying an efficient representation of the *wh*-dependency input
- results: a learner adopting this representation strategy successfully attains the target knowledge
- discussion and future direction

child directed *wh*-dependencies

[1]Perkins & Lidz 2021; [2]De Villiers et al. 2008;
[3]Pearl and Bates 2022; [4]Pearl and Sprouse 2013

child directed *wh*-dependencies

- Learning period: 18 months to 4 years[1,2]

[1]Perkins & Lidz 2021; [2]De Villiers et al. 2008;
[3]Pearl and Bates 2022; [4]Pearl and Sprouse 2013

child directed *wh*-dependencies

- Learning period: 18 months to 4 years[1,2]
- Number of dependencies estimation:[3]

  waking hours X utterances per hour X wh-dep per utterance

  ≈ 2,146,324

[1]Perkins & Lidz 2021; [2]De Villiers et al. 2008;
[3]Pearl and Bates 2022; [4]Pearl and Sprouse 2013

child directed *wh*-dependencies

- Learning period: 18 months to 4 years[1,2]
- Number of dependencies estimation:[3]

  waking hours X utterances per hour X wh-dep per utterance

  $\approx$ 2,146,324
- extracted 12,704 *wh*-dependencies from the CHILDES Treebank[4]

[1]Perkins & Lidz 2021; [2]De Villiers et al. 2008;
[3]Pearl and Bates 2022; [4]Pearl and Sprouse 2013

model input

Pearl & Sprouse 2013

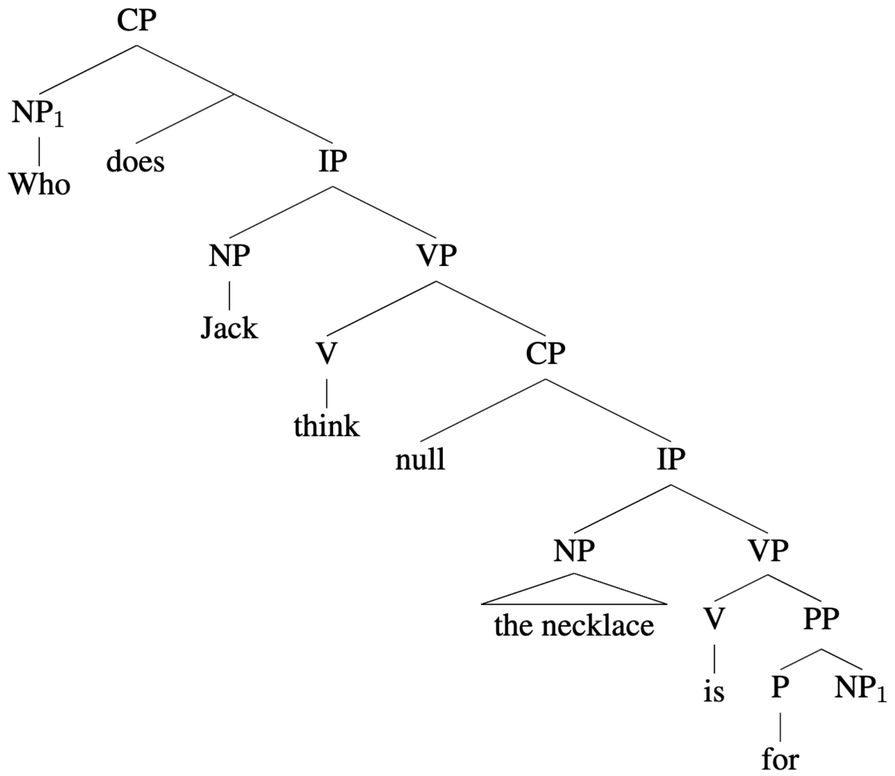
model input

- Following past work, we assume the relevant aspect of the wh-dependency tree is the dependency path

Pearl & Sprouse 2013

model input

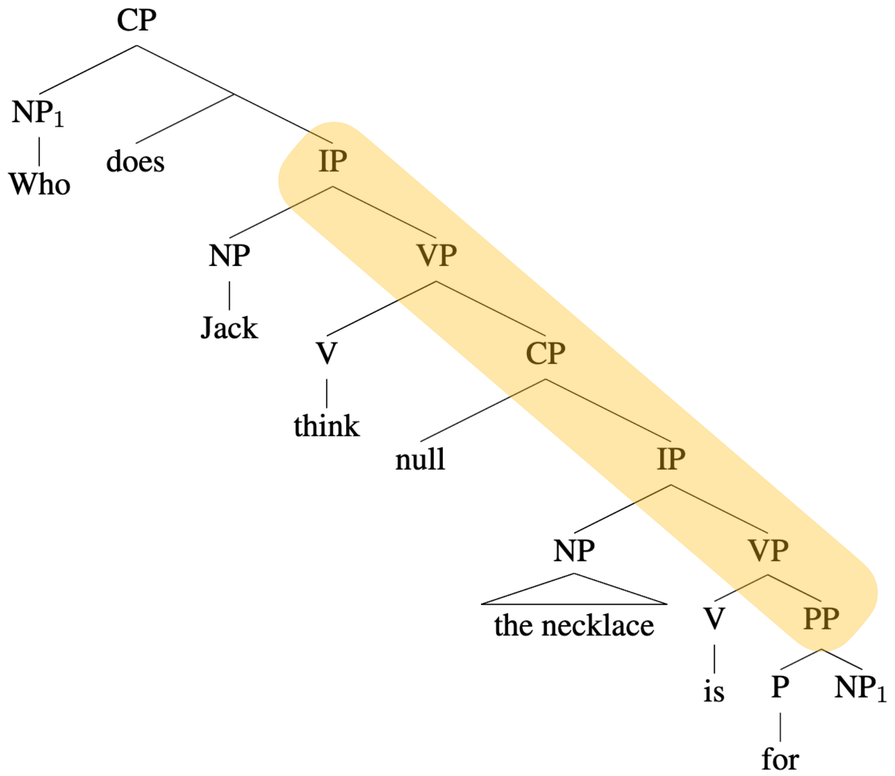
- Following past work, we assume the relevant aspect of the wh-dependency tree is the dependency path



Pearl & Sprouse 2013

model input

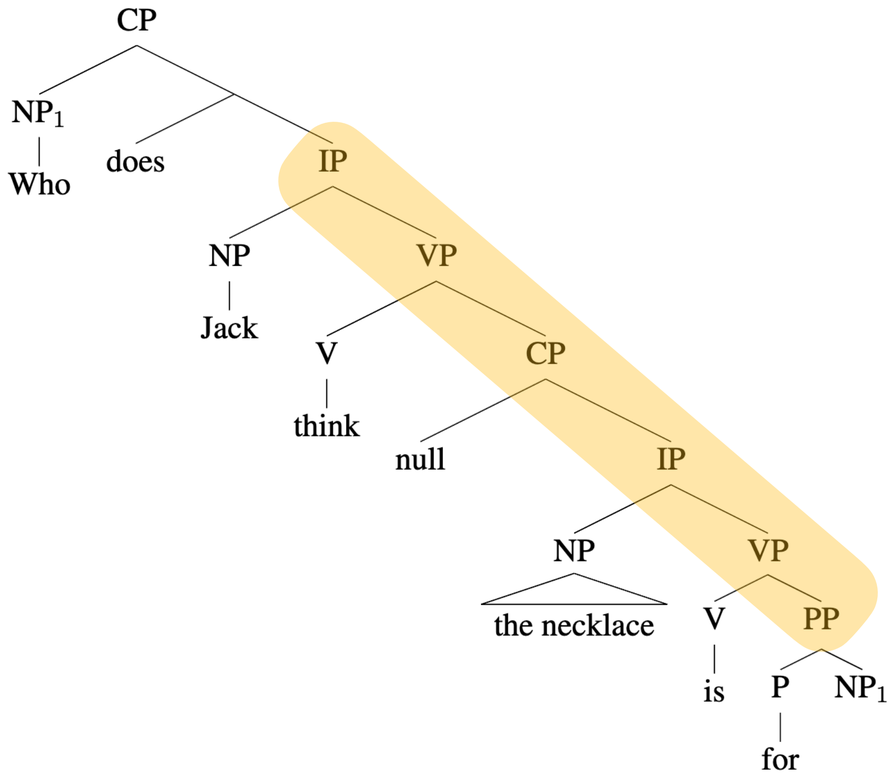
- Following past work, we assume the relevant aspect of the wh-dependency tree is the dependency path

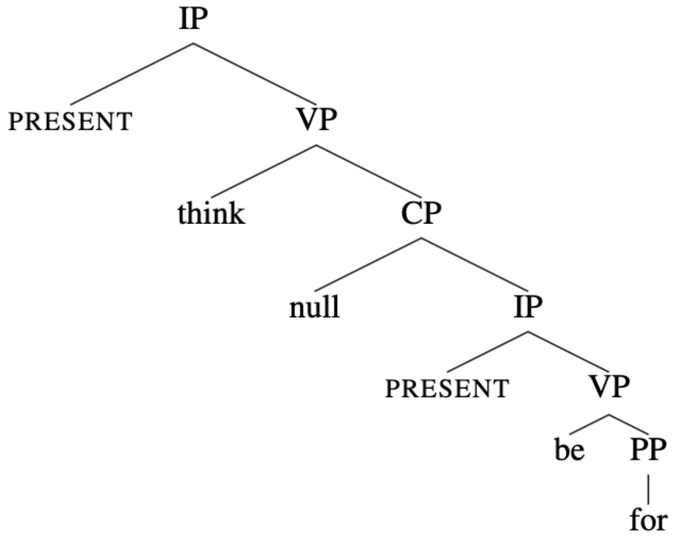


Pearl & Sprouse 2013

model input

- Following past work, we assume the relevant aspect of the wh-dependency tree is the dependency path





Pearl & Sprouse 2013

Position 4	Position 3	Position 2	Position 1
$\frac{1}{4^{0.1}} \approx 0.87$	$\frac{1}{3^{0.1}} \approx 0.90$	$\frac{1}{2^{0.1}} \approx 0.93$	$\frac{1}{1^{0.1}} = 1$
$\frac{1}{4^{2}} \approx 0.06$	$\frac{1}{3^{2}} \approx 0.11$	$\frac{1}{2^{2}} = 0.25$	$\frac{1}{1^{2}} = 1$

Position 4	Position 3	Position 2	Position 1
$IP_{PRESENT}$	VP_{think}	CP_{NULL}	$IP_{PRESENT}$

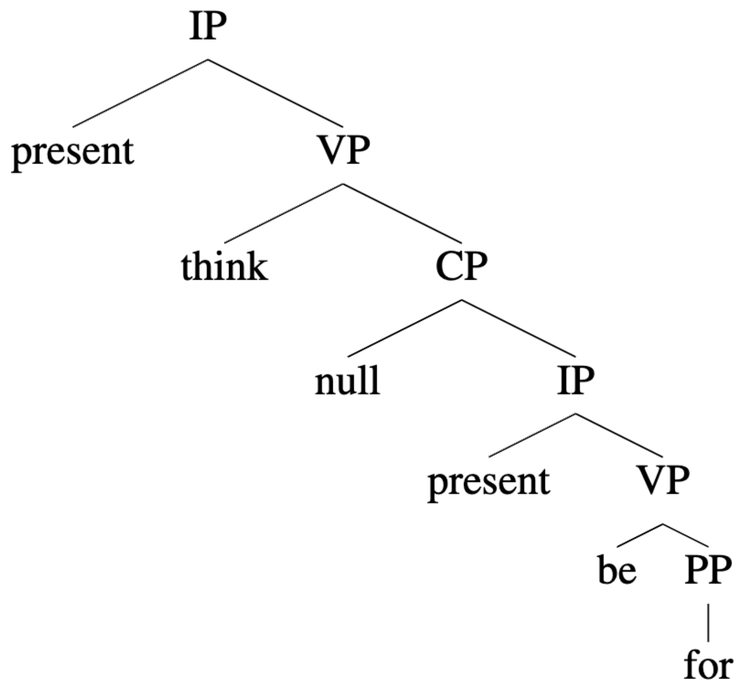
model input

Example *Wh*-Dependency	**Count**	**Percent** of **Stimuli**	**Cumulative Percent**
What's that?	3704	29.2%	29.2%
Who's that?	1502	11.8%	41.0%
What are you doing?	696	5.5%	46.5%
What did you do?	466	3.7%	50.1%
What was that?	264	2.1%	52.2%

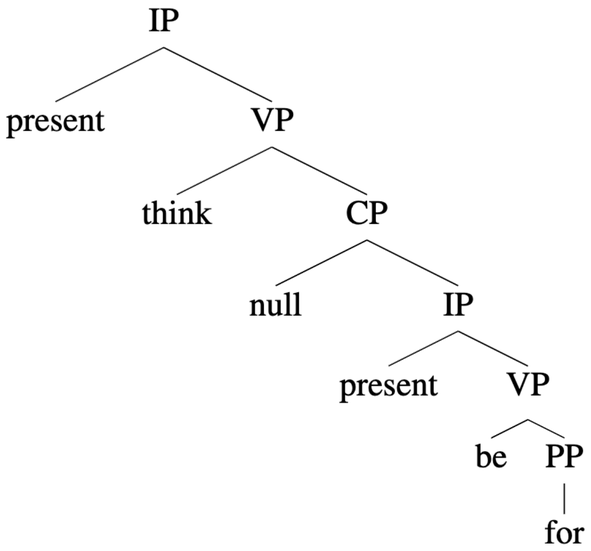
roadmap

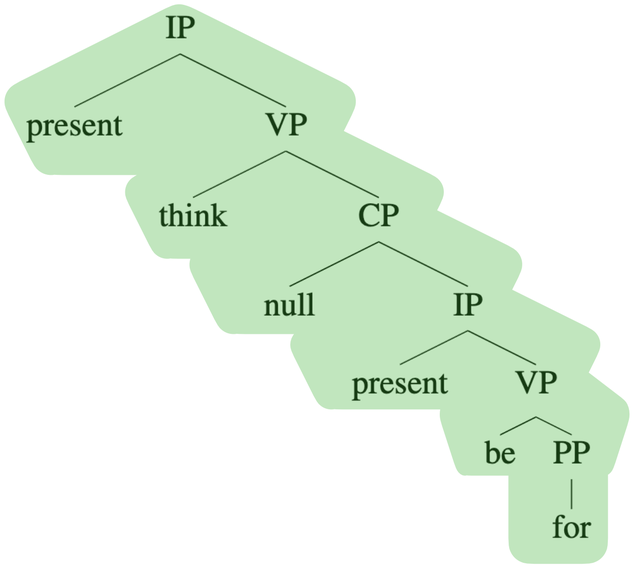
- output: behavioral patterns for these dependencies
- input: what do children hear and what input are we feeding our model
- learning objective: identifying an efficient representation of the *wh*-dependency input
- results: a learner adopting this representation strategy successfully attains the target knowledge
- discussion and future direction

learning theory

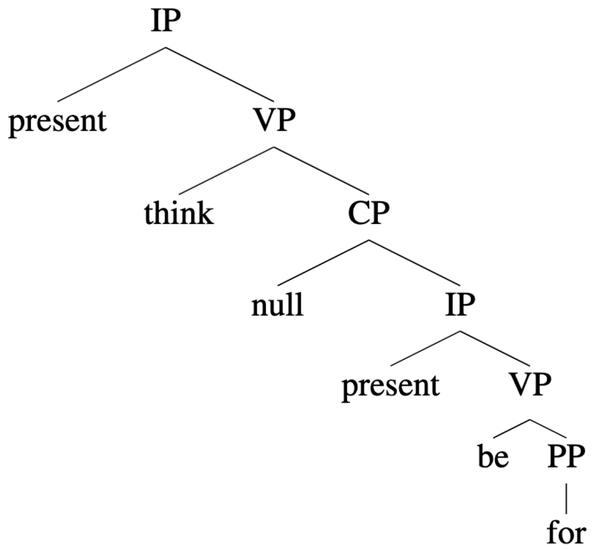


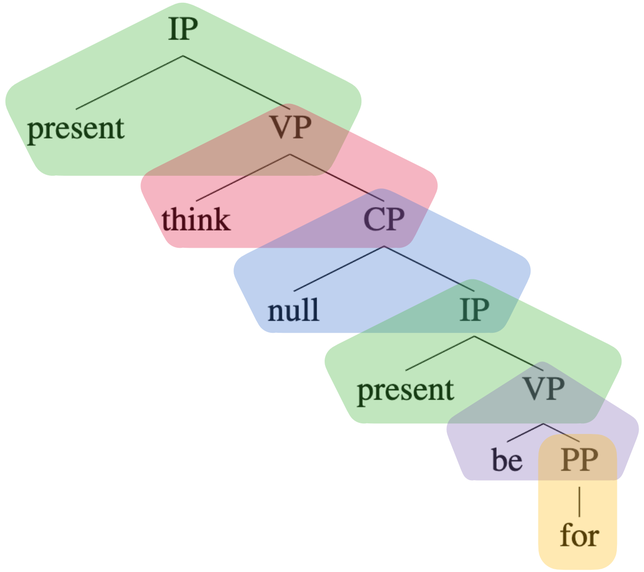
learning theory

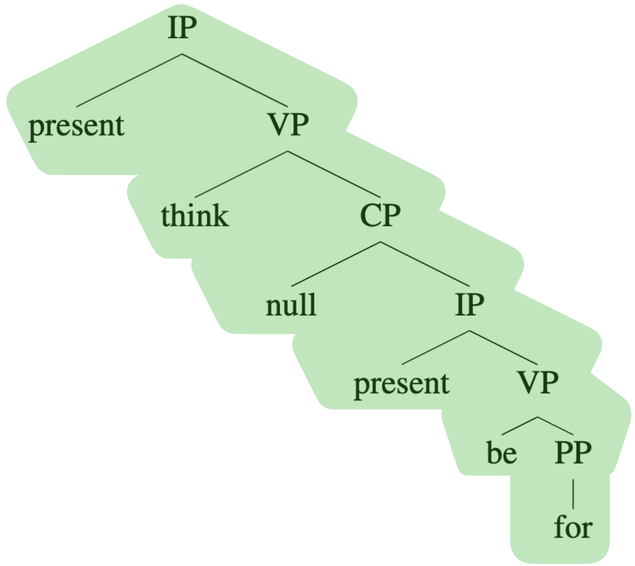




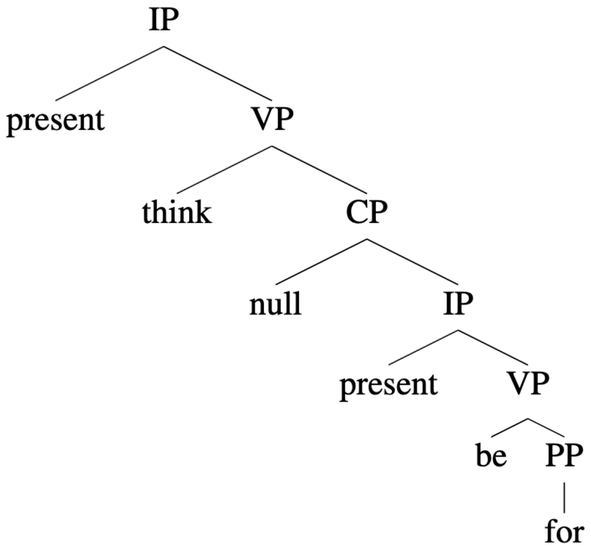
learning theory

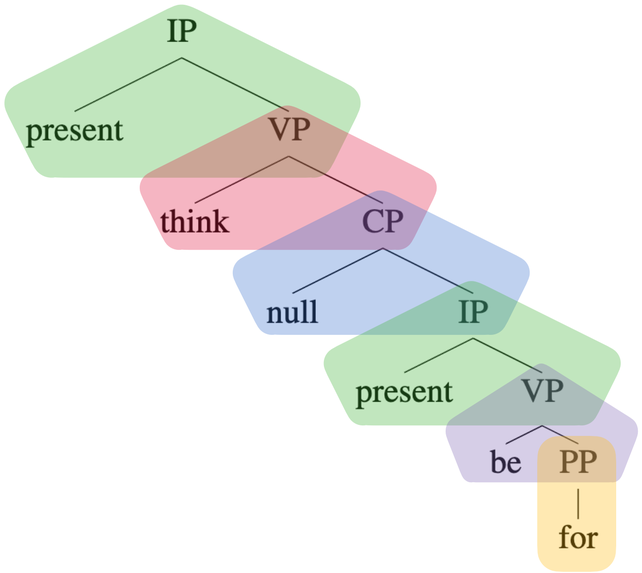


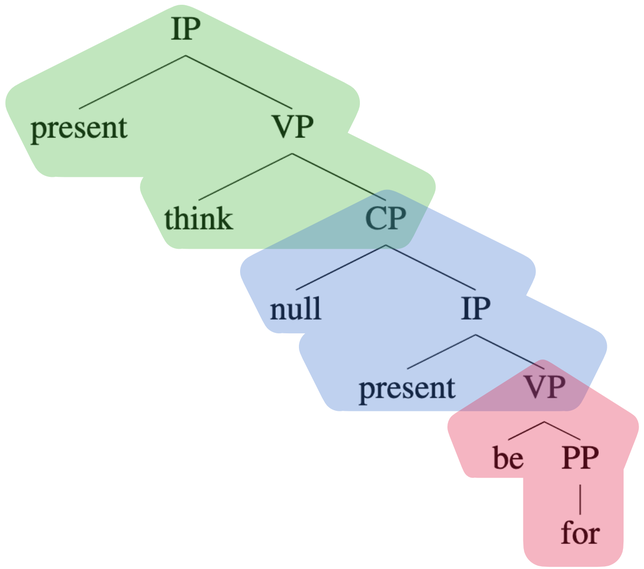


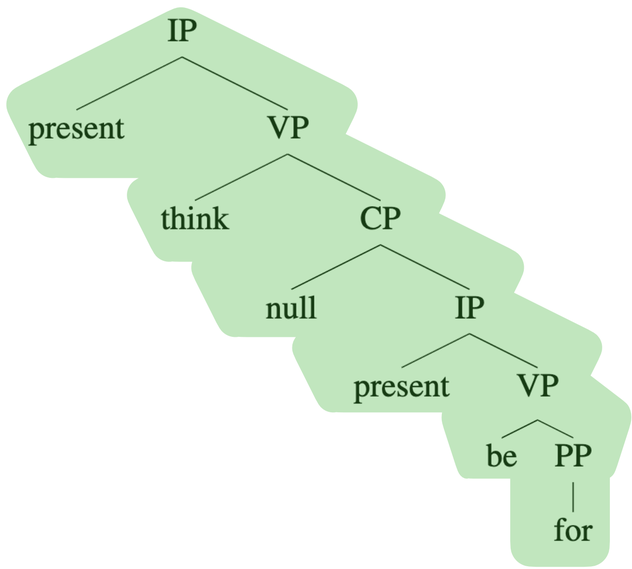


learning theory

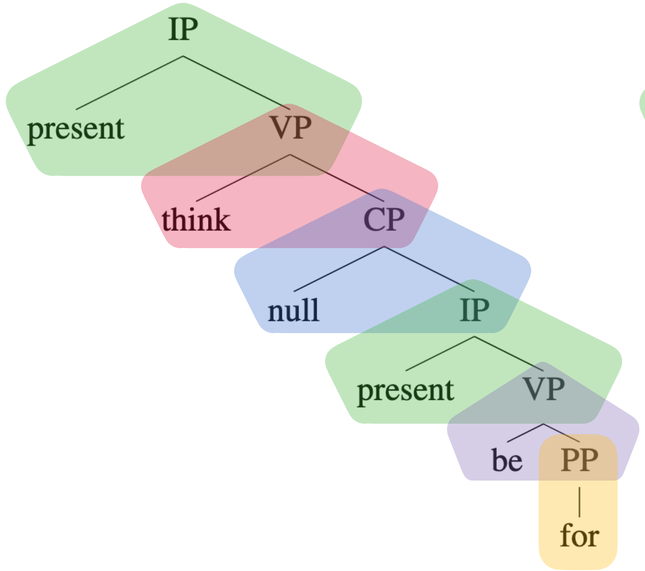




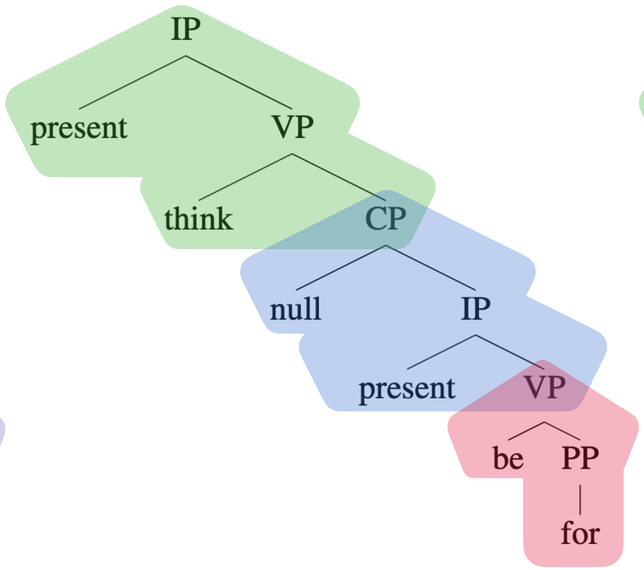




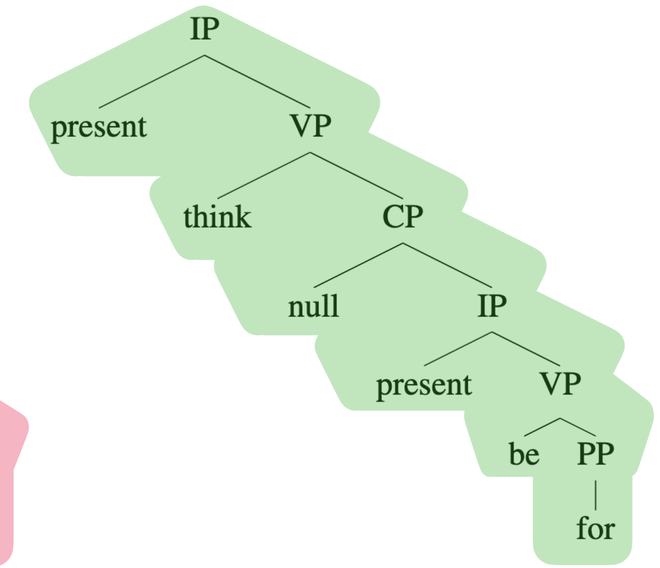
minimal learner



intermediate learner



maximal learner



Fragment Grammar

O'Donnell et al. 2009

Fragment Grammar

- A type of a PCFG

O'Donnell et al. 2009

Fragment Grammar

- A type of a PCFG
- Input of tree structures

O'Donnell et al. 2009

Fragment Grammar

- A type of a PCFG
- Input of tree structures
- Using bayesian inference to learn a representation of tree fragments and corresponding probabilities

O'Donnell et al. 2009

Fragment Grammar

- A type of a PCFG
- Input of tree structures
- Using bayesian inference to learn a representation of tree fragments and corresponding probabilities
- Using this representation, we can test novel stimuli

O'Donnell et al. 2009

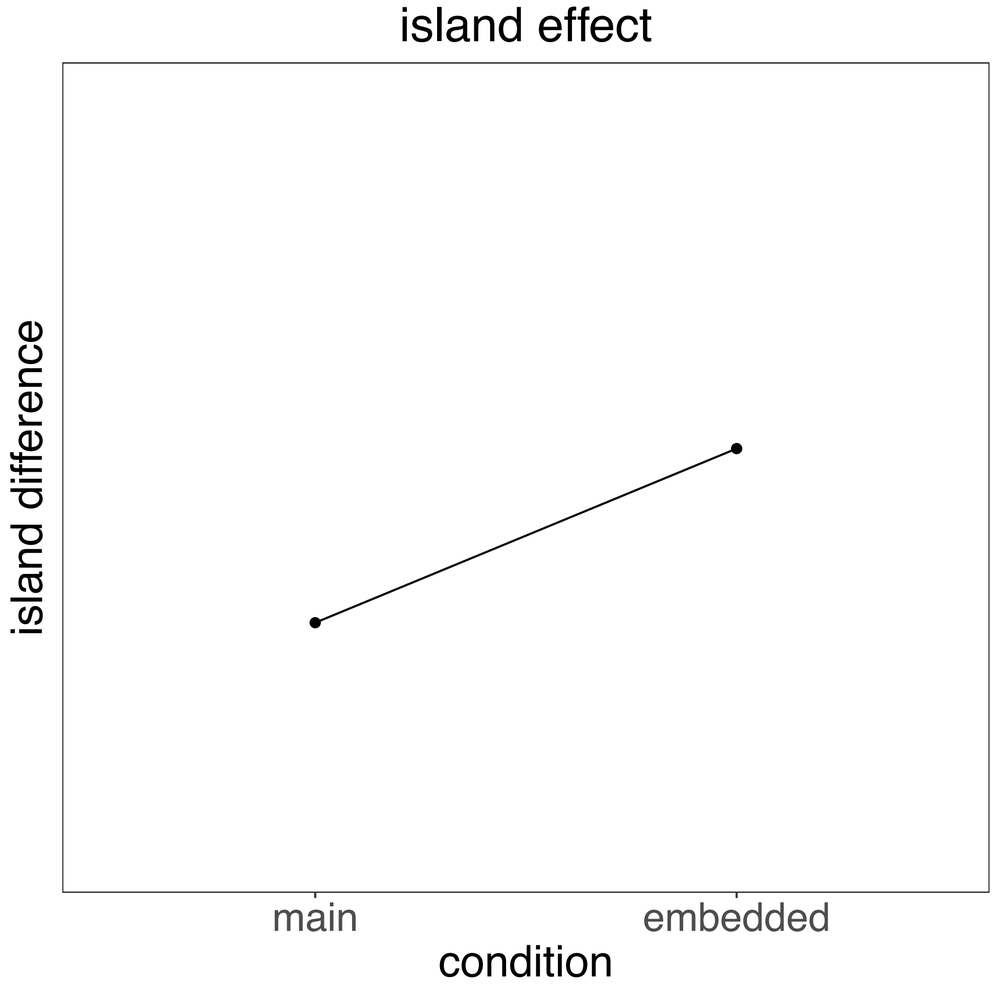
testing FG representation

- linking average log probability of the pieces with acceptability

roadmap

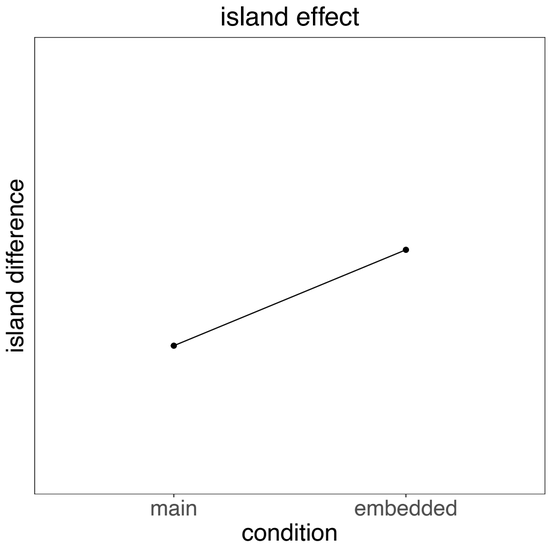
- output: behavioral patterns for these dependencies
- input: what do children hear and what input are we feeding our model
- learning objective: identifying an efficient representation of the *wh*-dependency input
- results: a learner adopting this representation strategy successfully attains the target knowledge
- discussion and future direction

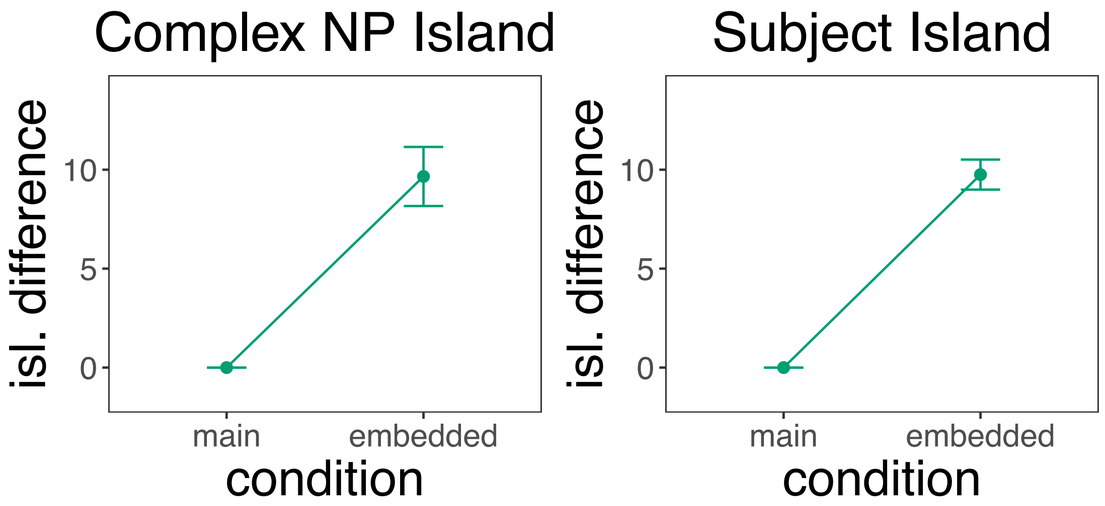
results

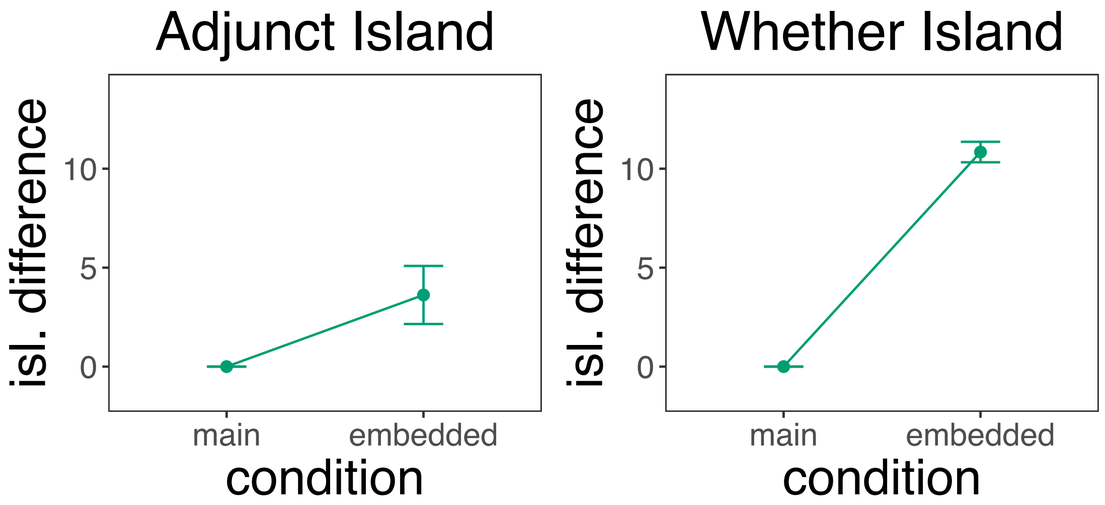


Sprouse et al. 2012

results

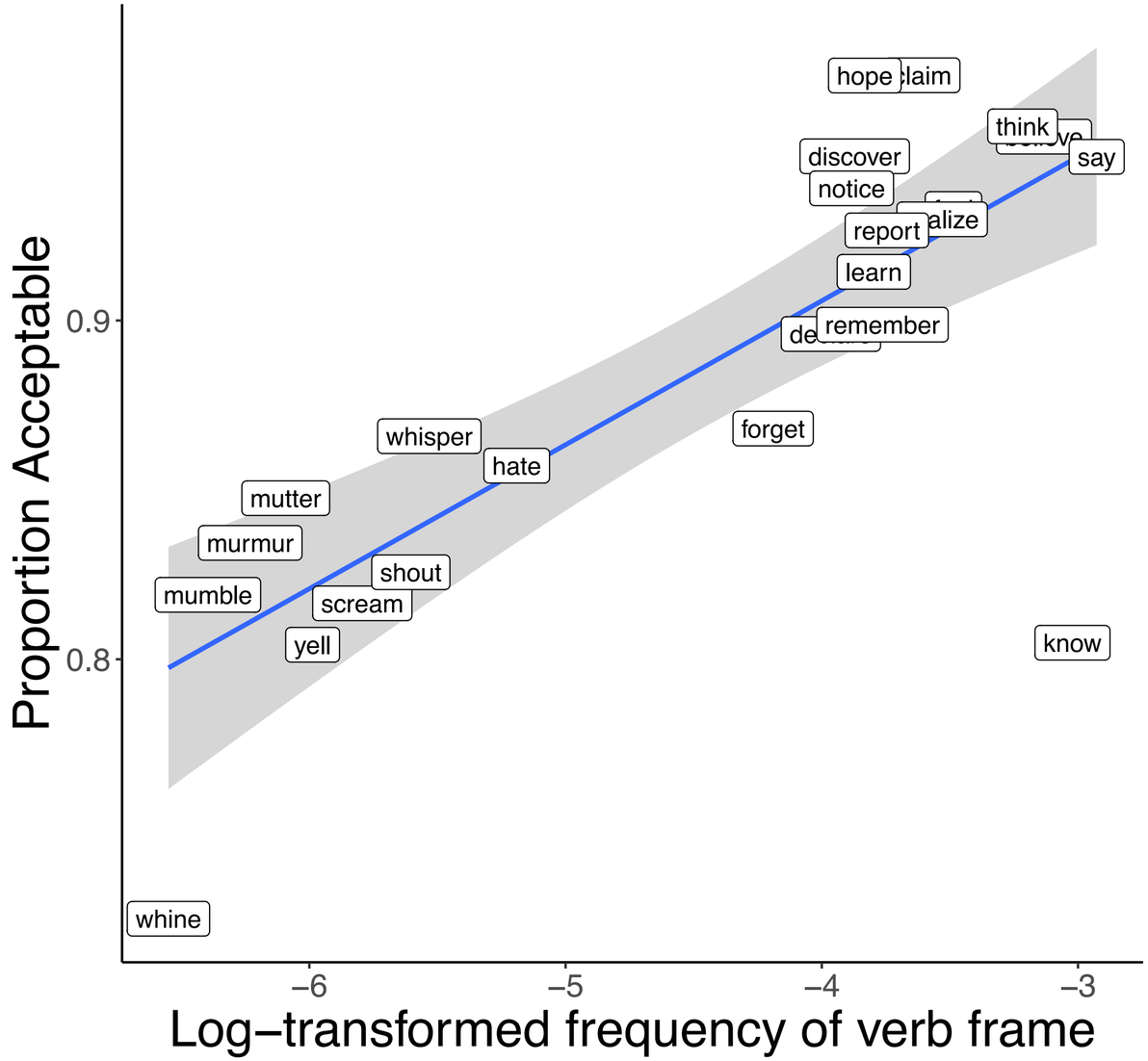






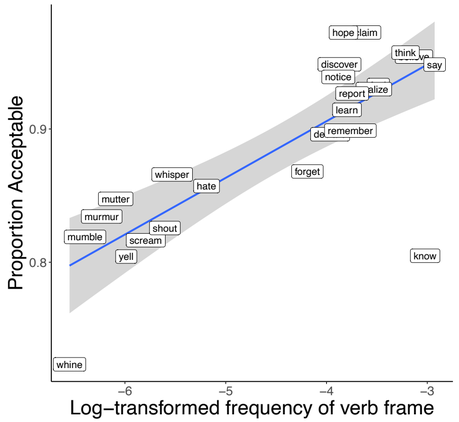
Sprouse et al. 2012

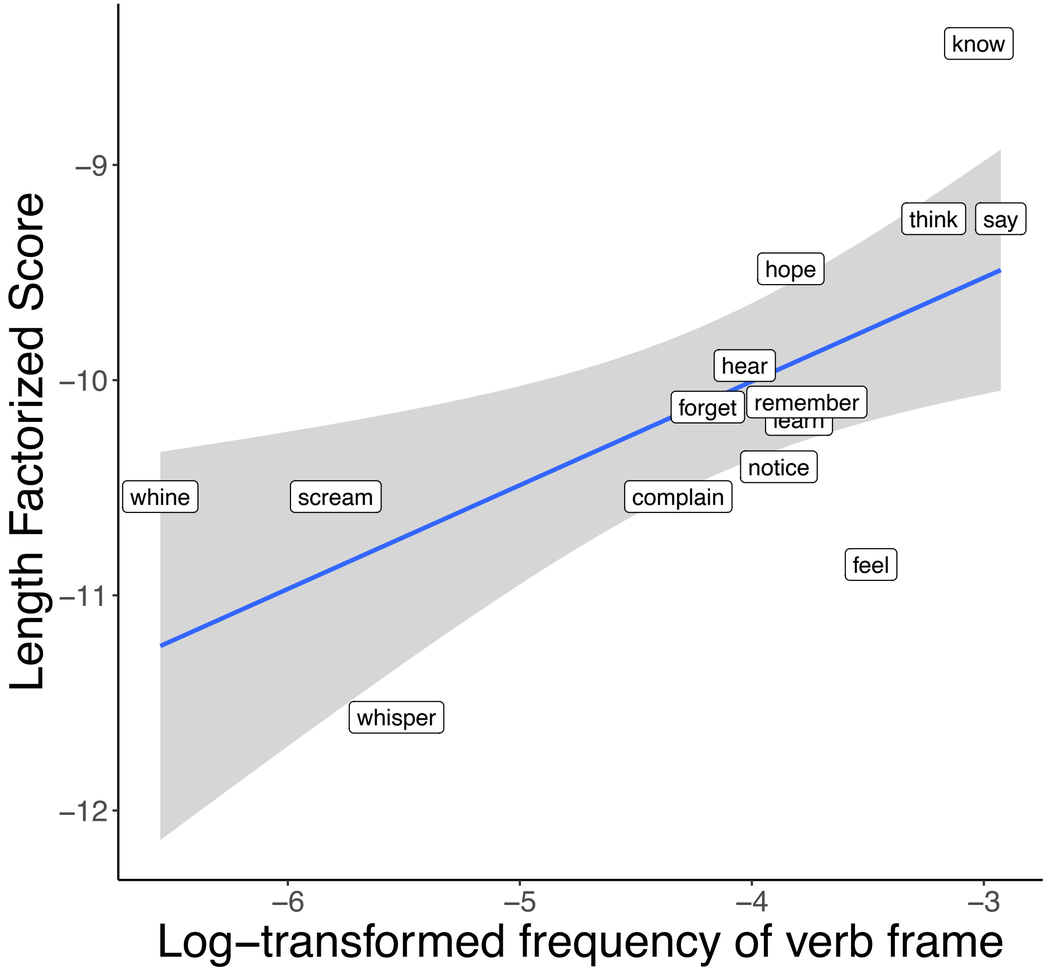
results



Liu et al. 2022

results





Liu et al. 2022

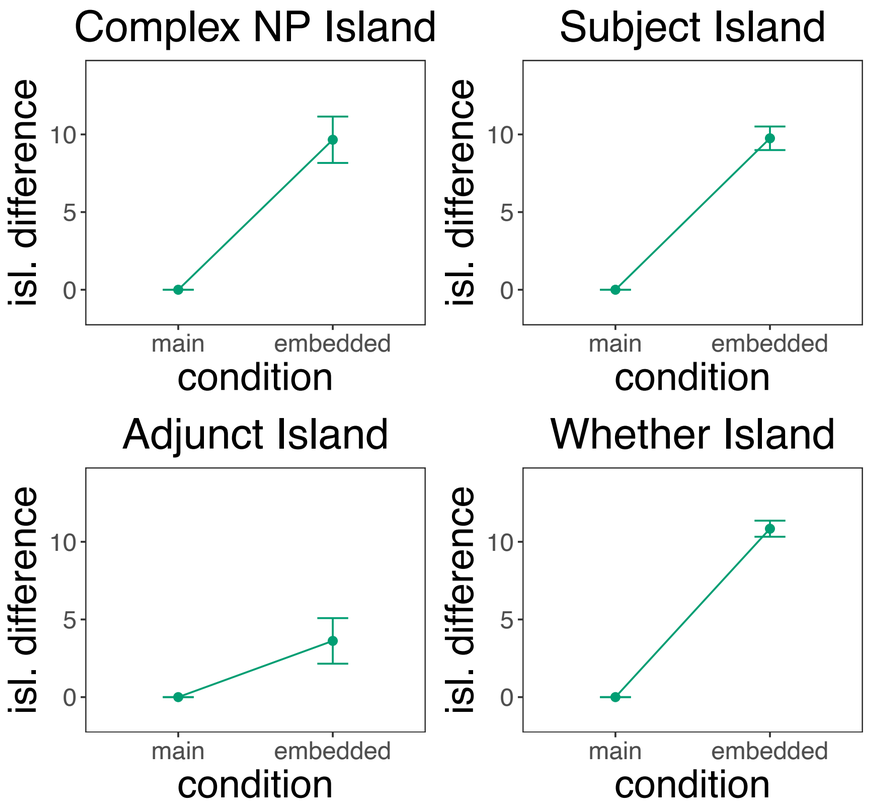
results

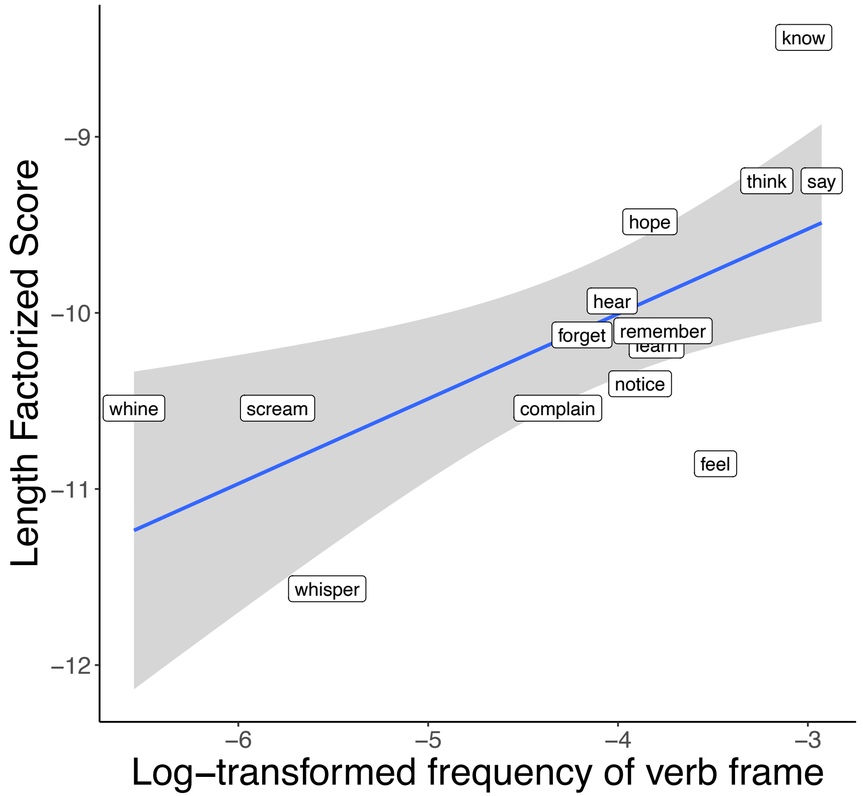
Utterance	How often children preferred the longer *wh*-dependency	Longer dependency preferred
How did the boy say he hurt himself?	0.80	✓
What did the mother say she bought?	0.79	✓
Who did the police woman help to call?	0.48	✗
Who did the little sister ask how to see?	0.25	✗
How did the boy who sneezed drink the milk?	0.20	✗
What did the boy fix the cat that was lying on the table with?	0.09	✗
How did the girl ask where to ride?	0.04	✗
Who did the boy ask what to bring?	0.04	✗
How did the mom learn what to bake?	0.03	✗

results

Utterance	How often children preferred the longer *wh*-dependency	Longer dependency preferred	FG Prediction
How did the boy say he hurt himself?	0.80	✓	✓ (0.58)
What did the mother say she bought?	0.79	✓	✓ (0.61)
Who did the police woman help to call?	0.48	✗	✓ (0.55)
Who did the little sister ask how to see?	0.25	✗	✗ (0.00)
How did the boy who sneezed drink the milk?	0.20	✗	✗ (0.00)
What did the boy fix the cat that was lying on the table with?	0.09	✗	✗ (0.00)
How did the girl ask where to ride?	0.04	✗	✗ (0.00)
Who did the boy ask what to bring?	0.04	✗	✗ (0.00)
How did the mom learn what to bake?	0.03	✗	✗ (0.00)

results



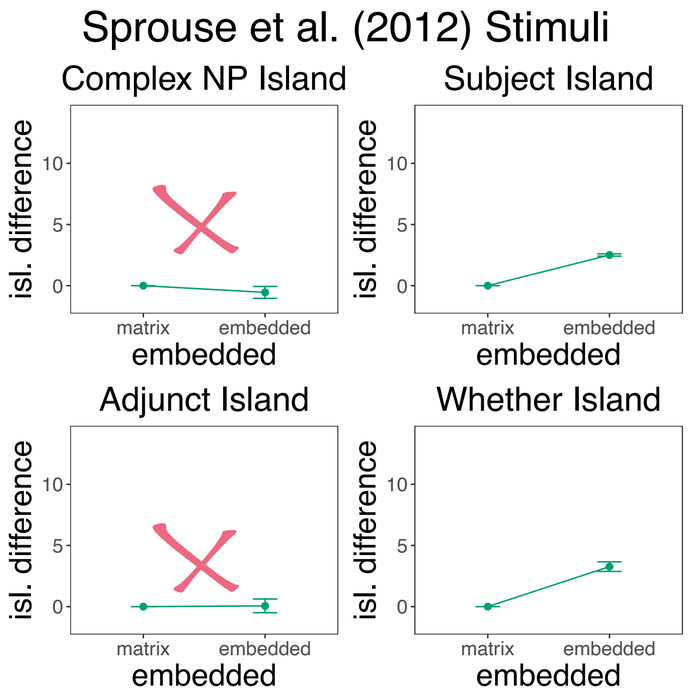


Longer dependency preferred	FG Prediction
✓	✓ (0.58)
✓	✓ (0.61)
✗	✓ (0.55)
✗	✗ (0.00)
✗	✗ (0.00)
✗	✗ (0.00)
✗	✗ (0.00)
✗	✗ (0.00)
✗	✗ (0.00)

fixed sized chunks baseline

fixed sized chunks baseline

Fully Lexicalized Trigram



Longer dependency preferred	Model Prediction
✓	✗
✓	✗
✗	✓
✗	✗
✗	✗
✗	✗
✗	✗
✗	✗
✗	✗

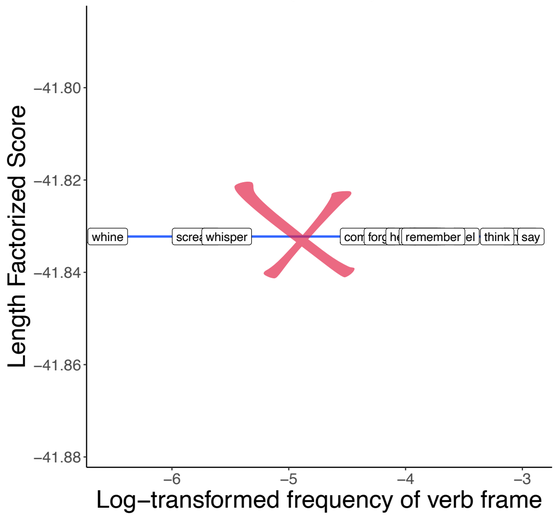
Lexicalized Main Verb Trigram

Longer dependency preferred	Model Prediction
✓	✗
✓	✗
✗	✓
✗	✗
✗	✗
✗	✓
✗	✗
✗	✗
✗	✗

fixed sized chunks baseline

Lexicalized Complementizer Trigram

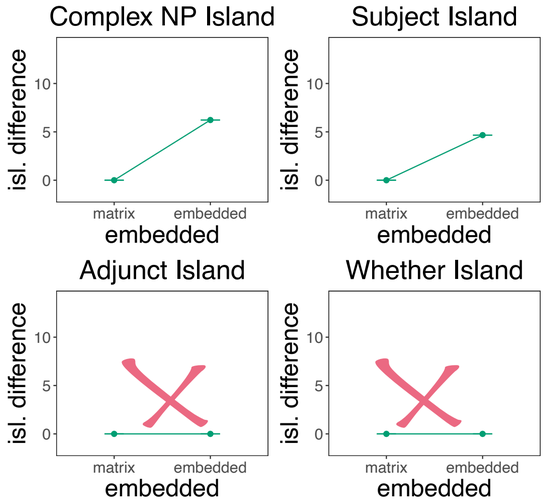
Liu et al. (2022) Stimuli



Longer dependency preferred	Model Prediction
✓	✗
✓	✗
✗	✗
✗	✗
✗	✗
✗	✗
✗	✗
✗	✗
✗	✗

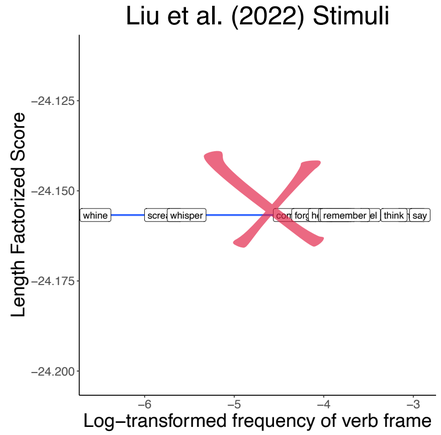
Phrasal Trigram

Sprouse et al. (2012) Stimuli



Longer dependency preferred	Model Prediction
✓	✗
✓	✗
✗	✗
✗	✗
✗	✗
✗	✗
✗	✗
✗	✗
✗	✗

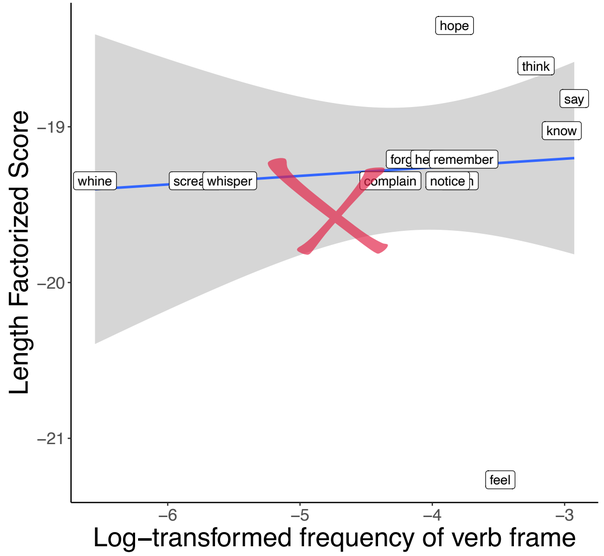
Liu et al. (2022) Stimuli



variable sized chunks baseline

Adaptor Grammar

Liu et al. (2022) Stimuli



Longer dependency preferred	Model Prediction
✓	✗
✓	✗
✗	✓
✗	✗
✗	✗
✗	✗
✗	✗
✗	✗
✗	✗

PCFG

Longer dependency preferred	Model Prediction
✓	✗
✓	✗
✗	✓
✗	✗
✗	✗
✗	✗
✗	✗
✗	✗
✗	✗

summary

summary

- We've seen that a modeled learner can acquire adult like *wh*-dependency knowledge by efficiently representing the input

summary

- We've seen that a modeled learner can acquire adult like *wh*-dependency knowledge by efficiently representing the input
  - these constraints comes for free with the goal of efficient representation

future directions

future directions

- Making changes to the model input

future directions

- Making changes to the model input
  - dense child directed corpora will better approximate the child’s input

future directions

- Making changes to the model input
  - dense child directed corpora will better approximate the child's input
  - give the model full trees

future directions

- Making changes to the model input
  - dense child directed corpora will better approximate the child's input
  - give the model full trees
  - not assuming a perfect representation of the input

future directions

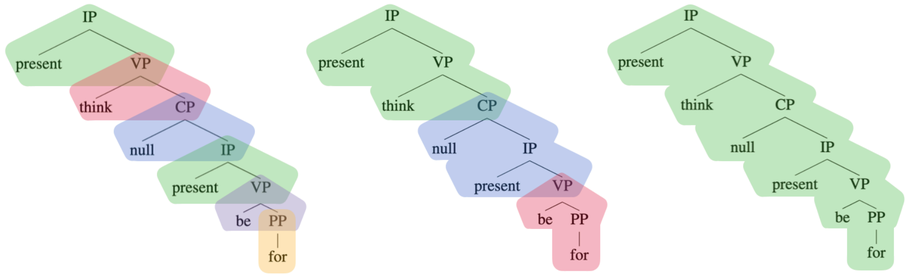
- Making changes to the model input
  - dense child directed corpora will better approximate the child's input
  - give the model full trees
  - not assuming a perfect representation of the input
- connecting this FG approach to other chunking literature

Acquiring *Wh*-Dependencies

Who does Jack think the necklace is for?

?

Efficient Representation

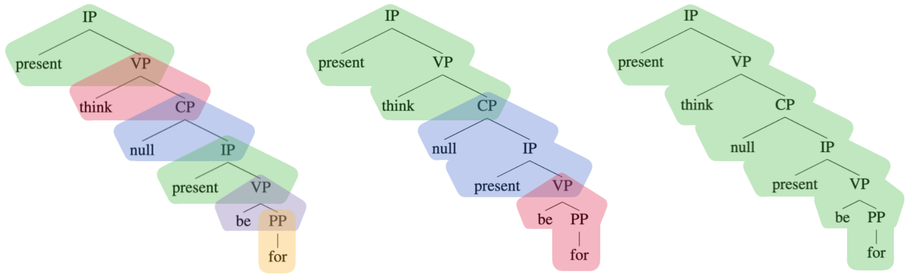


Acquiring *Wh*-Dependencies

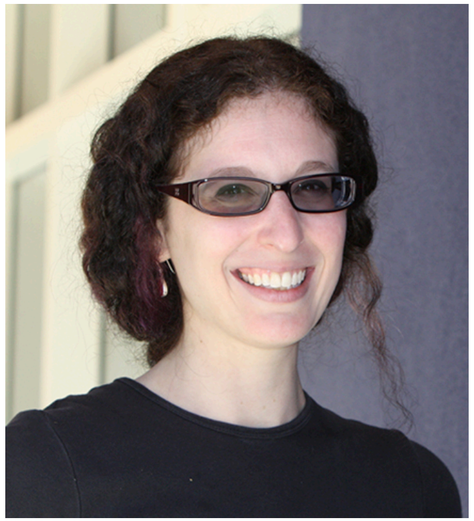
Who does Jack think the necklace is for?

✓

Efficient Representation



thank you!



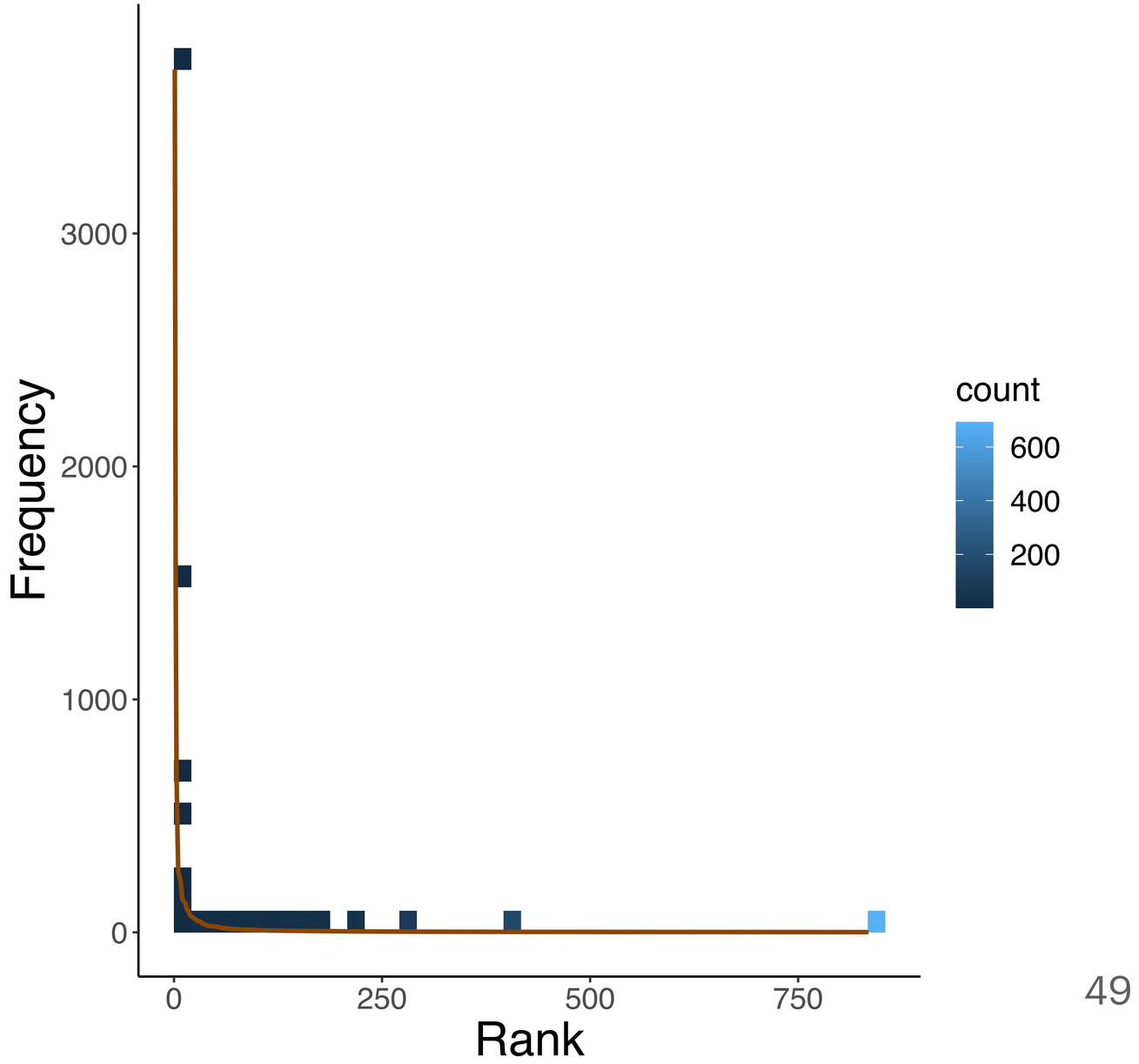
BU **Linguistics**
Conference on Language Development

SCiL SOCIETY FOR COMPUTATION IN LINGUISTICS

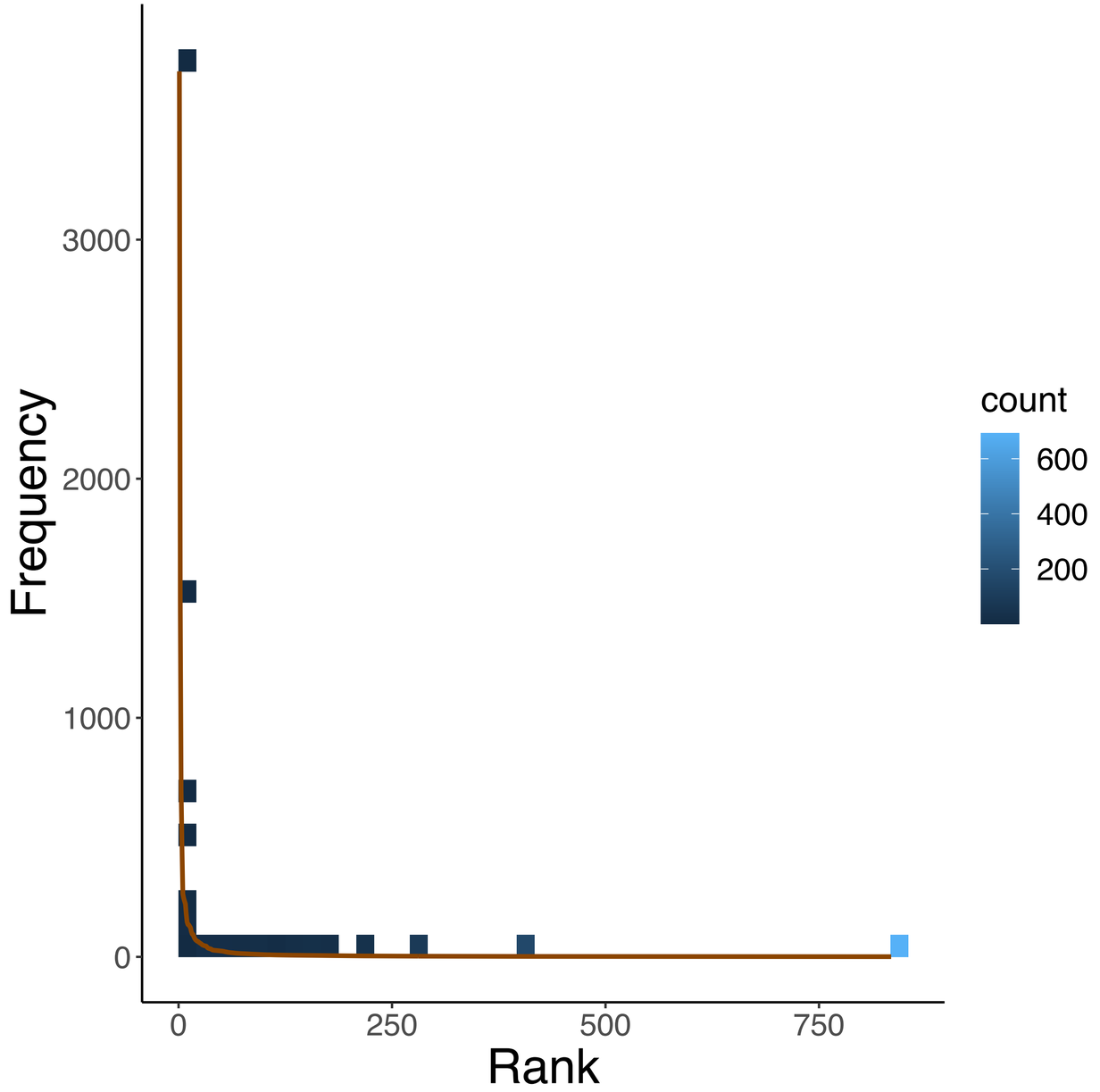
Department of Language Science
UCI School of Social Sciences

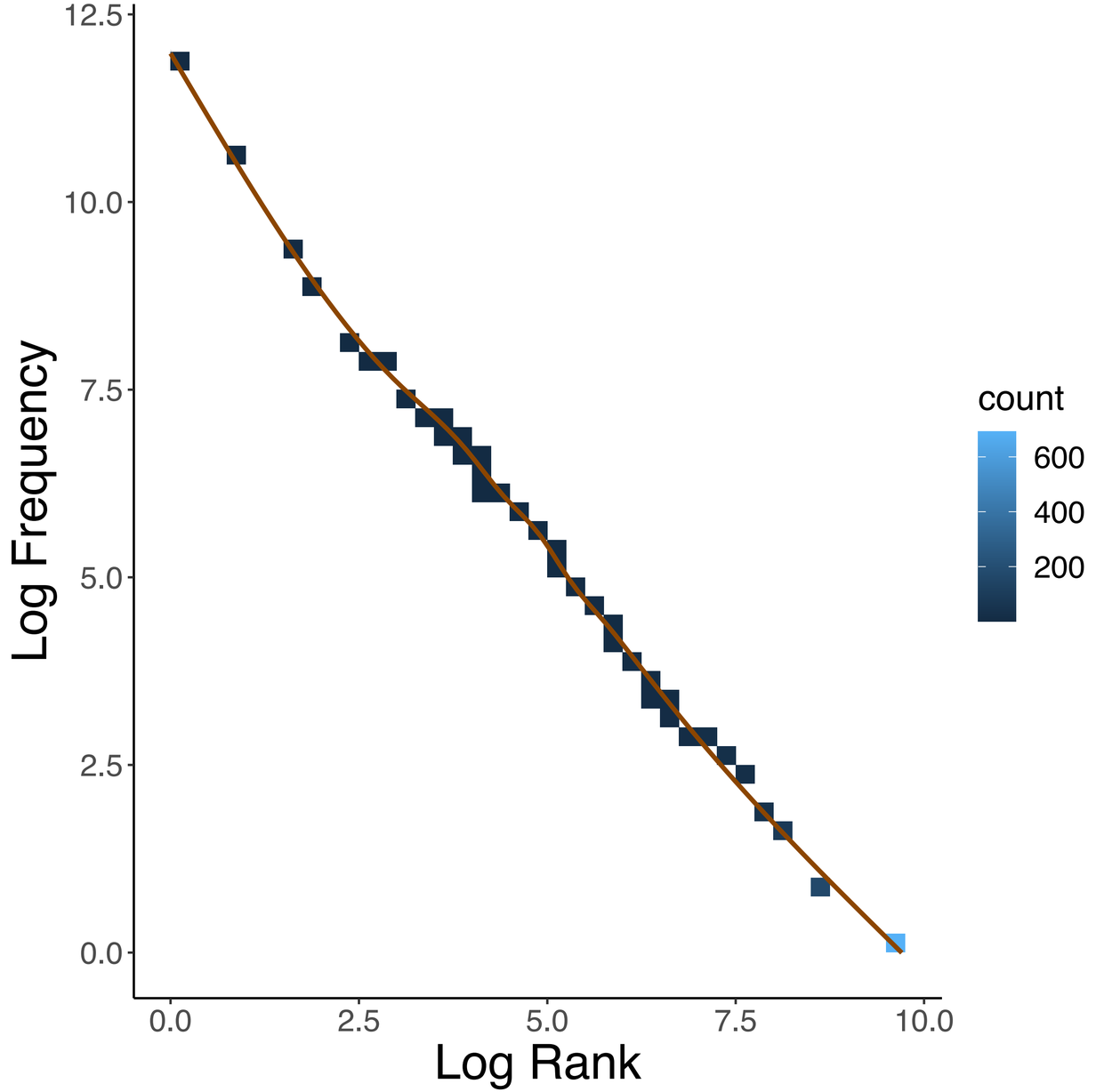
Zipfian Distribution

Zipfian Distribution



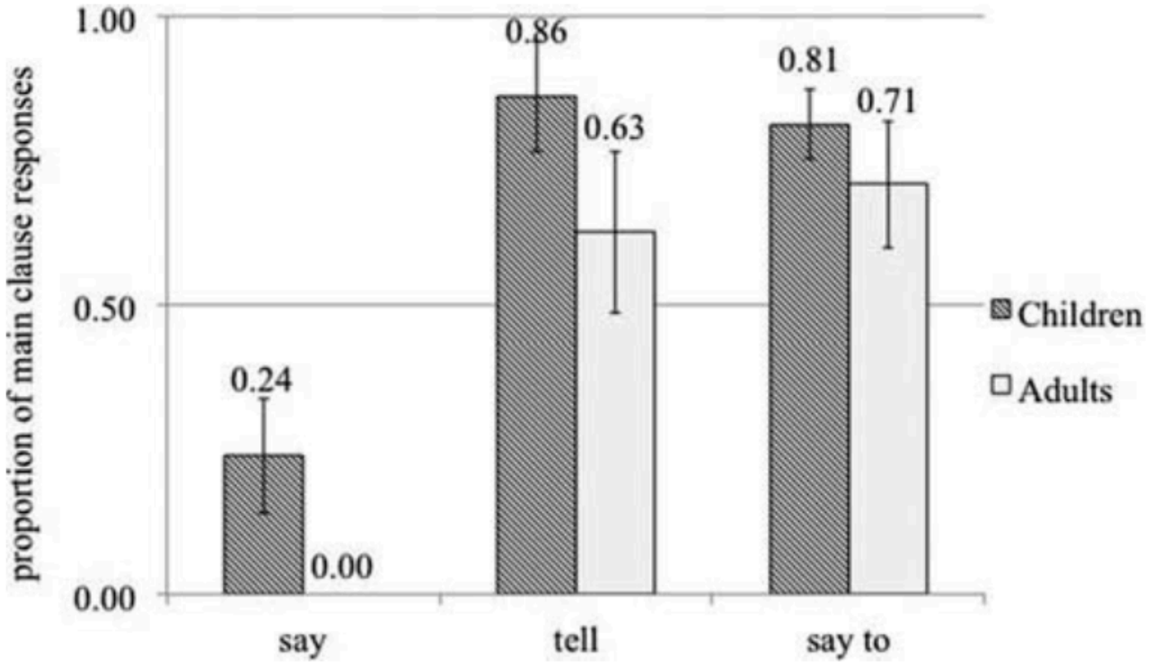
Zipfian Distribution





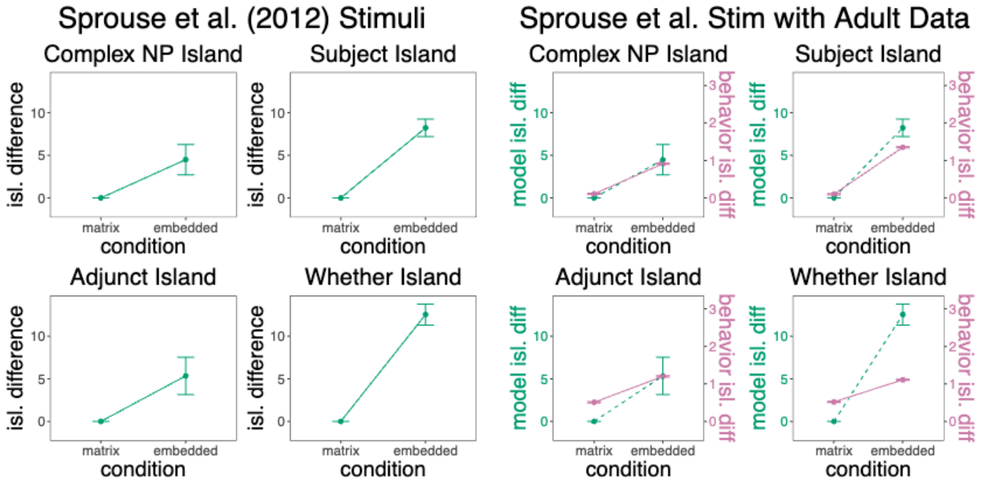
Omaki et al 2014 pattern not captured by syntactic path

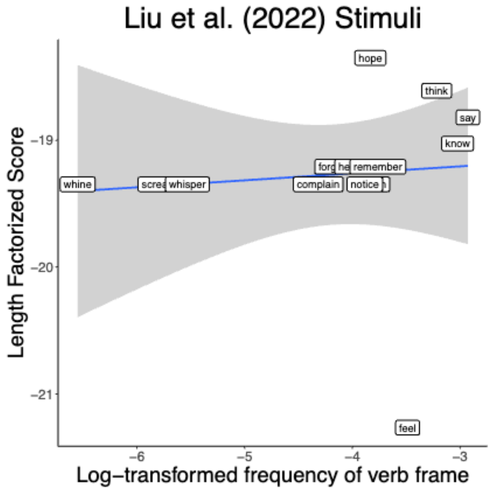
(6) Where did Lizzie {say | tell someone | say to someone} that she was gonna catch butterflies?

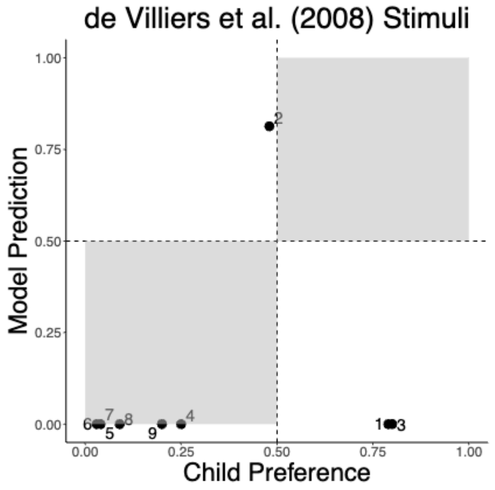


comparison results

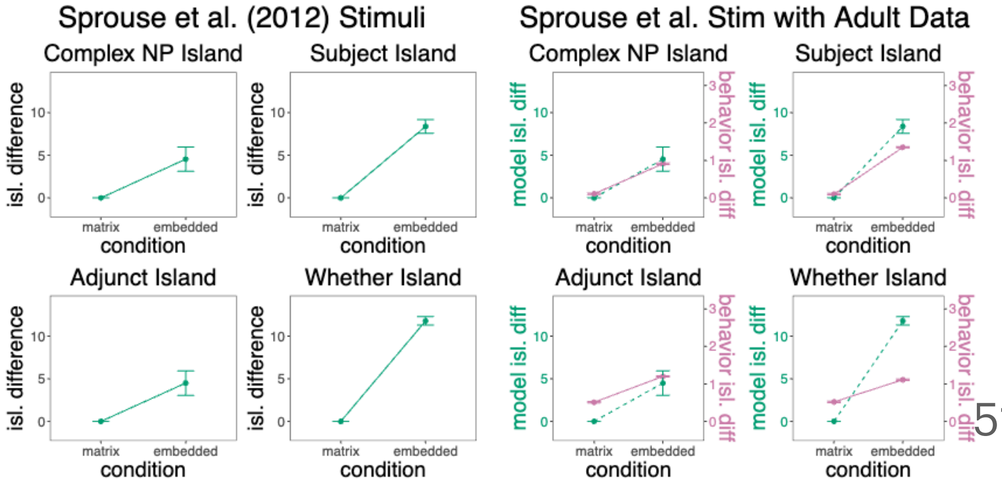
Adaptor Grammar Results

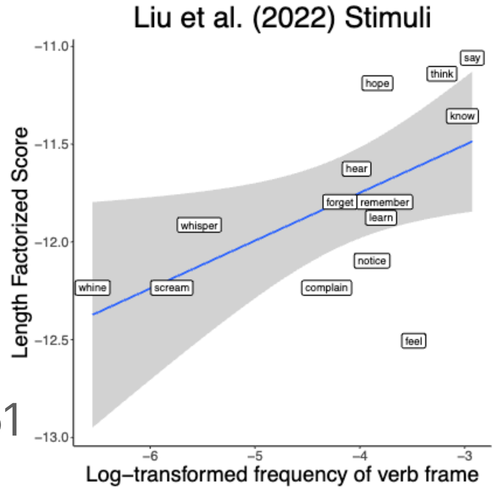


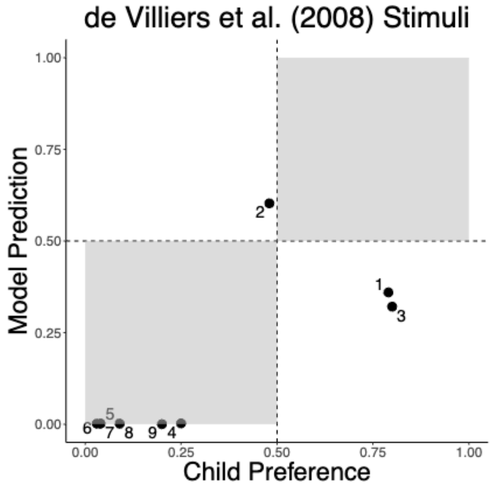




PCFG Results

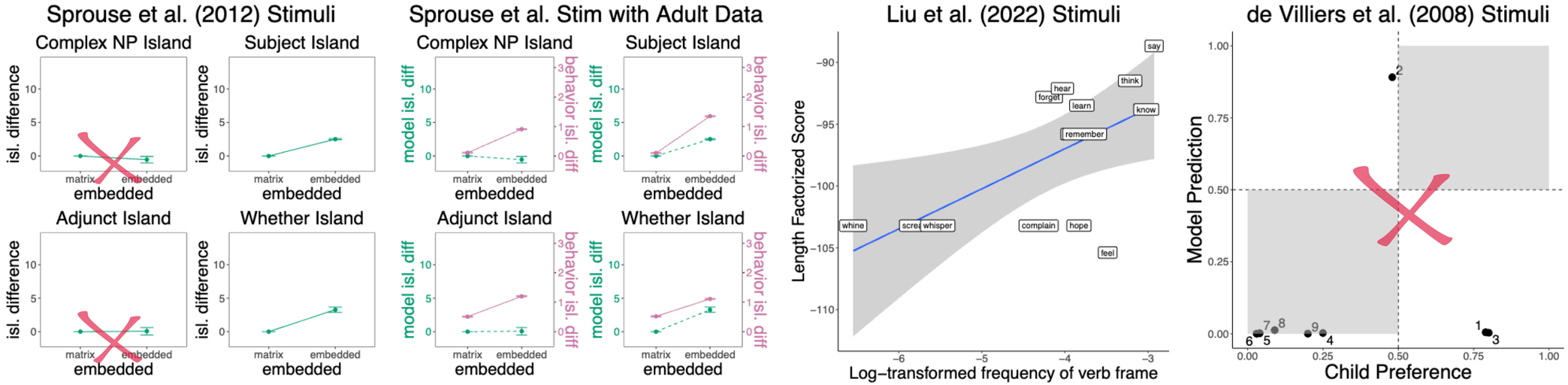




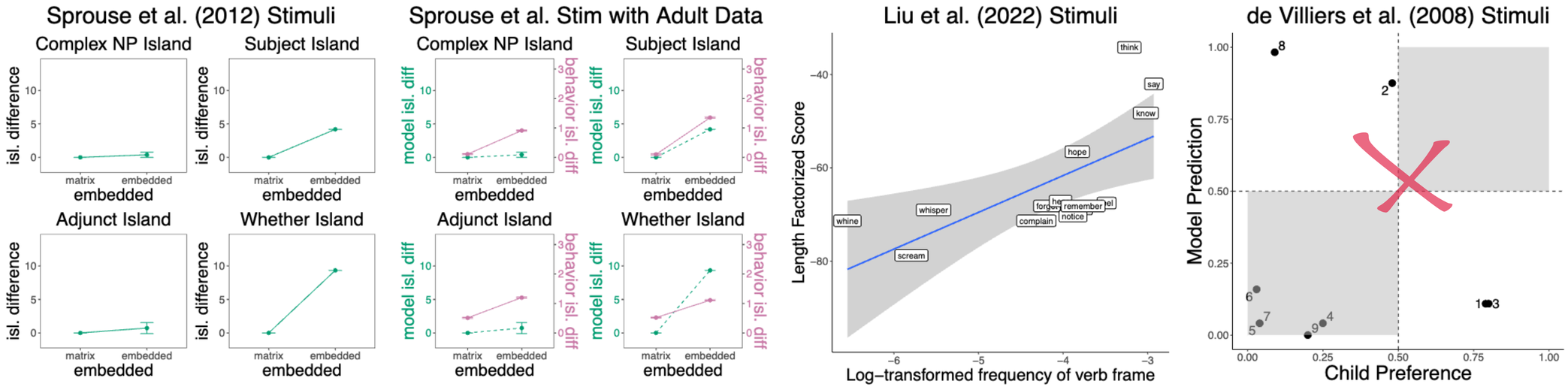


baseline results

Fully Lexicalized Trigram Results

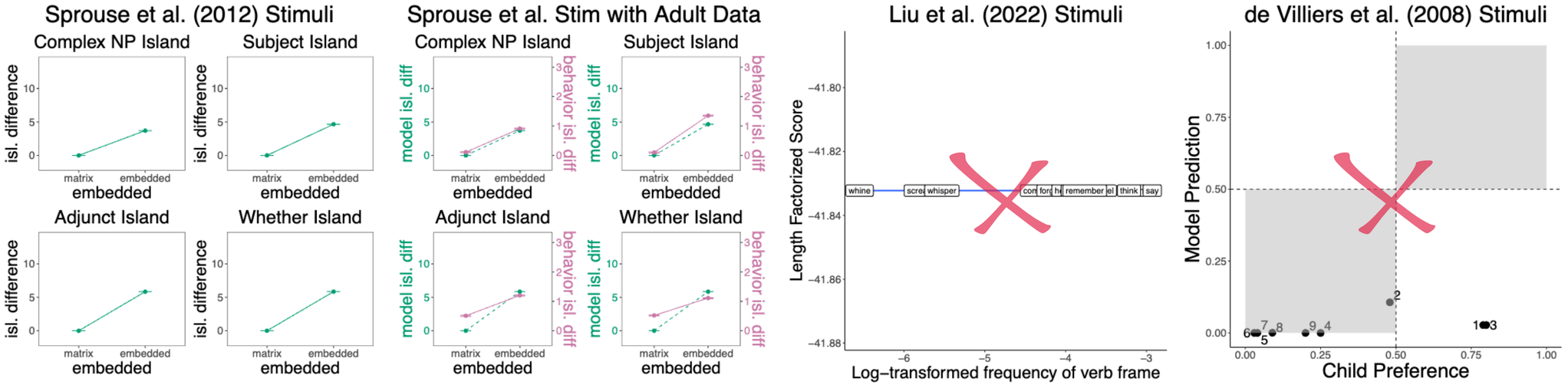


Lexicalized Main Verb Trigram Results



baseline results

Lexicalized Complimentizer Trigram Results



Phrasal Trigram Results