# What do RNN Language Models Learn about the Filler-Gap Dependency?


**Ethan Wilcox**
**Harvard University**

**Roger Levy**
**MIT**

**Takashi Morita**
**Kyoto University**

**Richard Futrell**
**UC Irvine**

# What do RNN Language Models Learn ___ about the Filler-Gap Dependency?


**Ethan Wilcox**
**Harvard University**

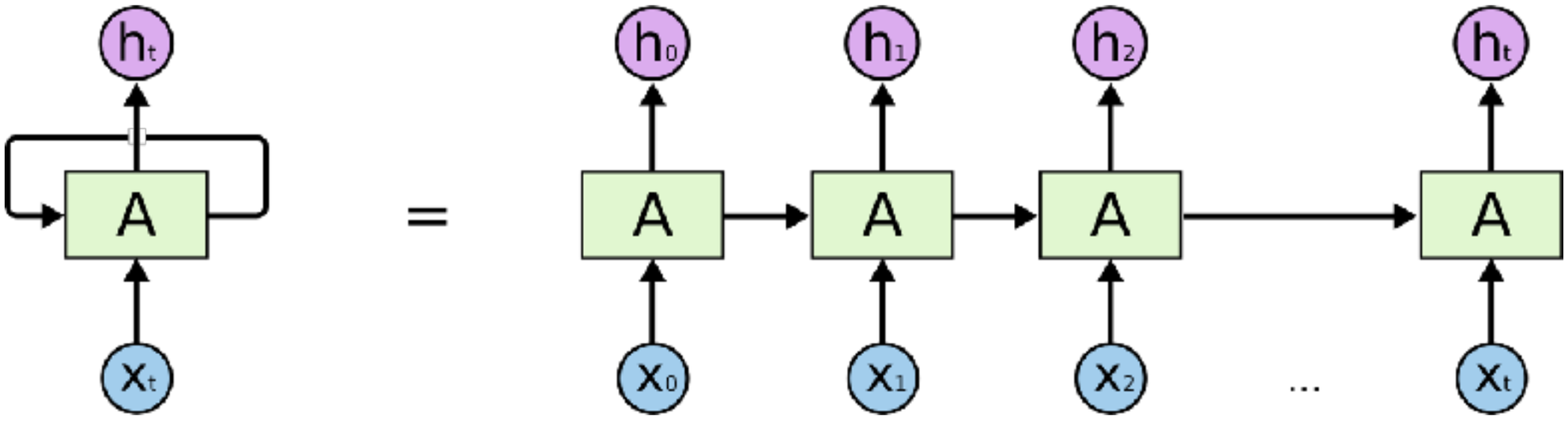
**Roger Levy**
**MIT**

**Takashi Morita**
**Kyoto University**

**Richard Futrell**
**UC Irvine**

# Objectives

- RNN LMs learn the **filler-gap dependency**.
- RNN LMs  learn some constraints on the dependency, known as "**island constraints.**"

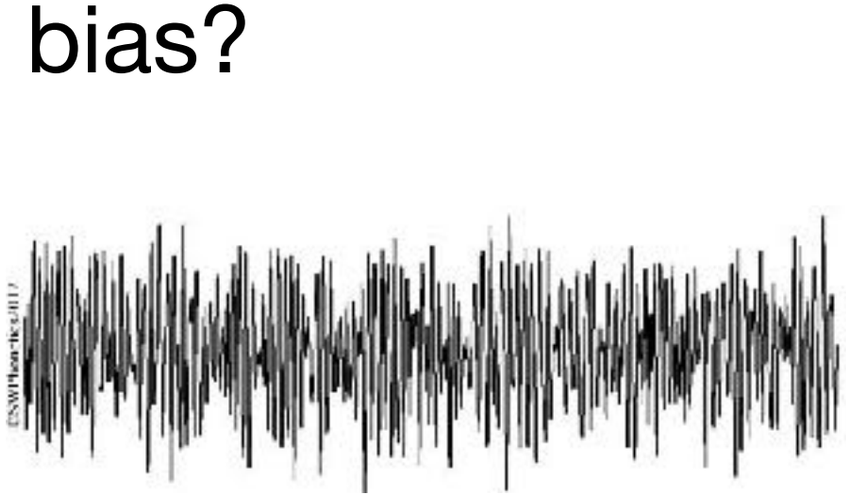


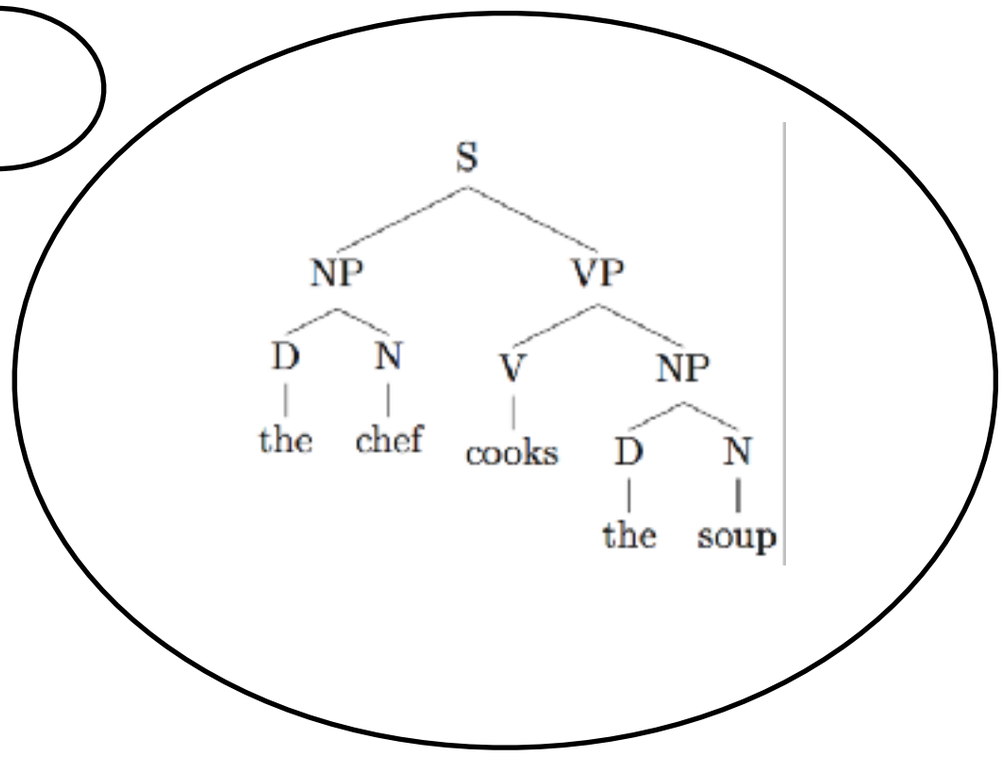
# Motivation

- **Technical Question:** What representations are these models learning and how are they learning them?
- **Theoretical Question:** What syntactic structures are easily learned by models without explicit hierarchical bias?

# Motivation

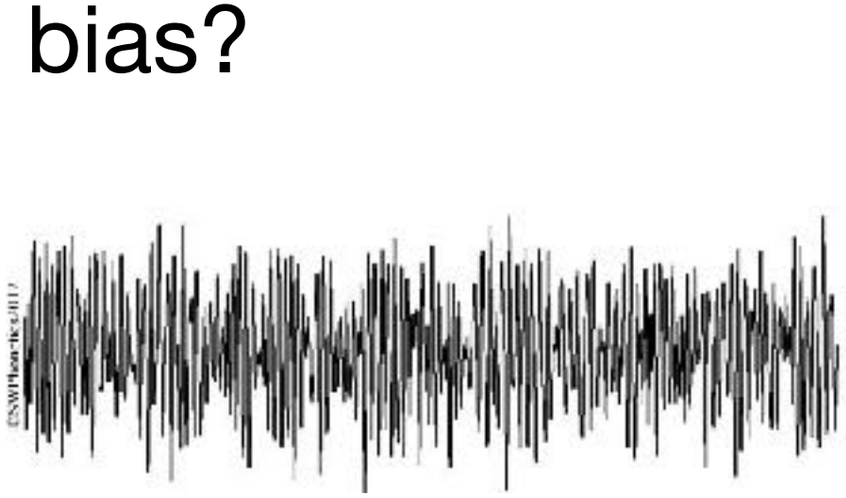
- **Technical Question:** What representations are these models learning and how are they learning them?
- **Theoretical Question:** What syntactic structures are easily learned by models without explicit hierarchical bias?

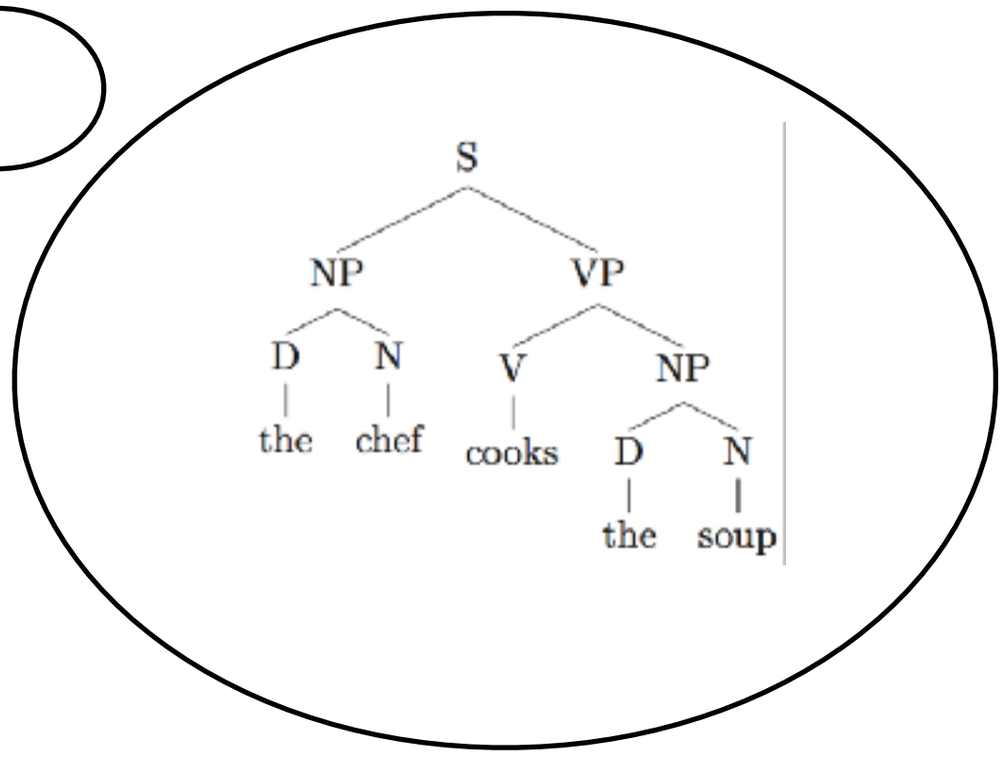




# Motivation

- **Technical Question:** What representations are these models learning and how are they learning them?
- **Theoretical Question:** What syntactic structures are easily learned by models without explicit hierarchical bias?






**UNIVERSAL CONSTRAINTS?**

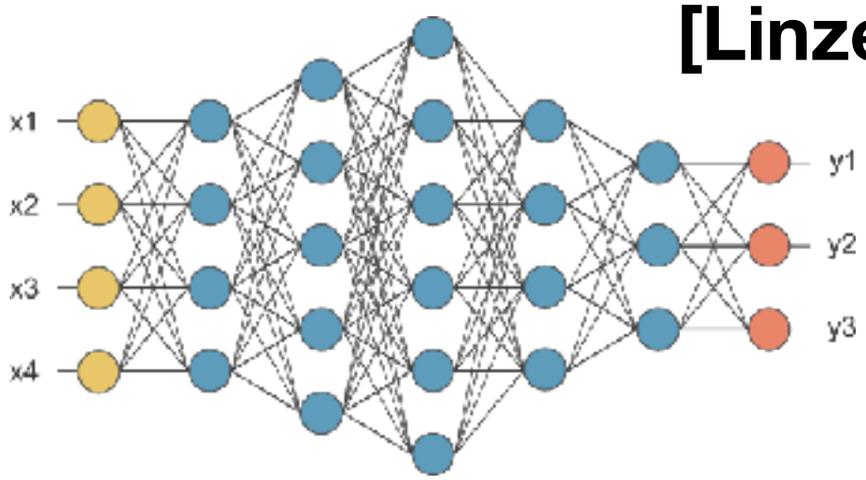
# Technical Motivation

## **Psycholinguistics Paradigm** for RNN Assessment

**(a) *“The keys to the cabinet is on the table”**

**(b) “The keys to the cabinet are on the table”**

**[Linzen et al. 2016]**

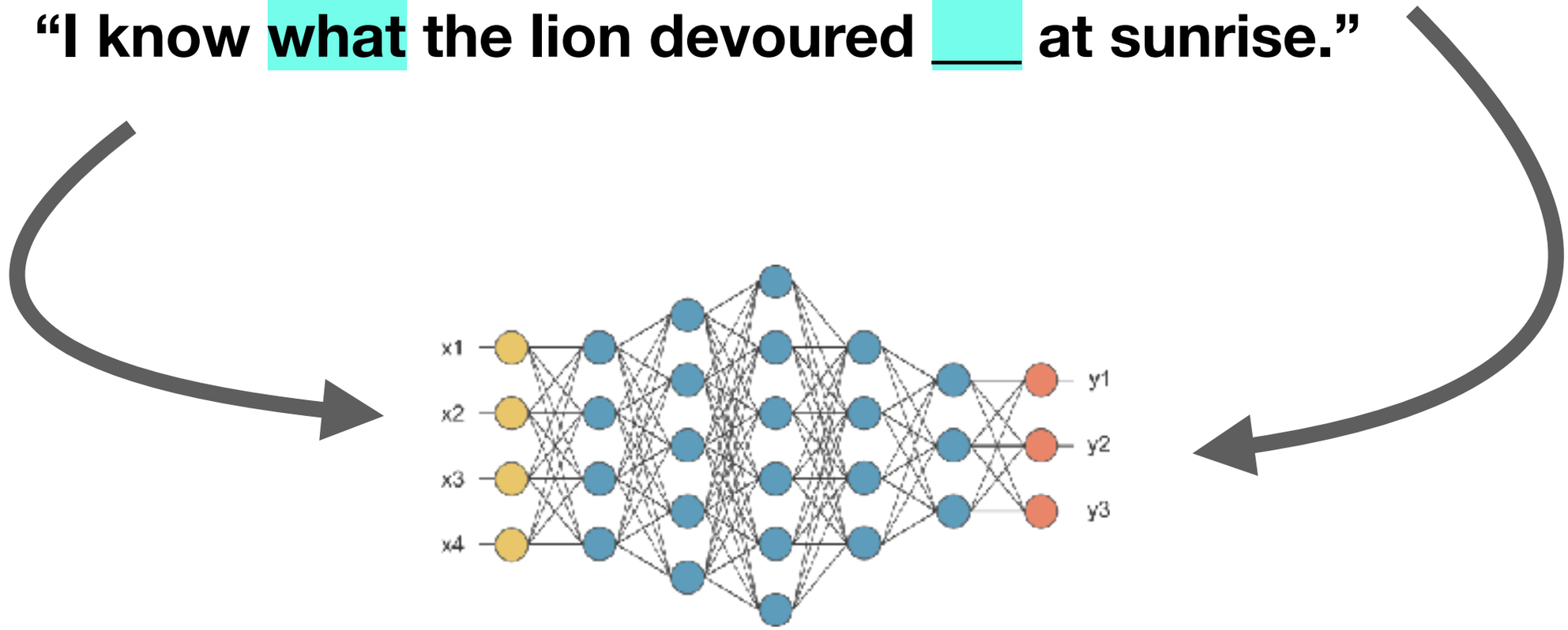



**(a) is SURPRISING!** **(b) is UNSURPRISING**

# Technical Motivation

## **Psycholinguistics Paradigm** for RNN Assessment

**“I know what the lion devoured ___ at sunrise.”**

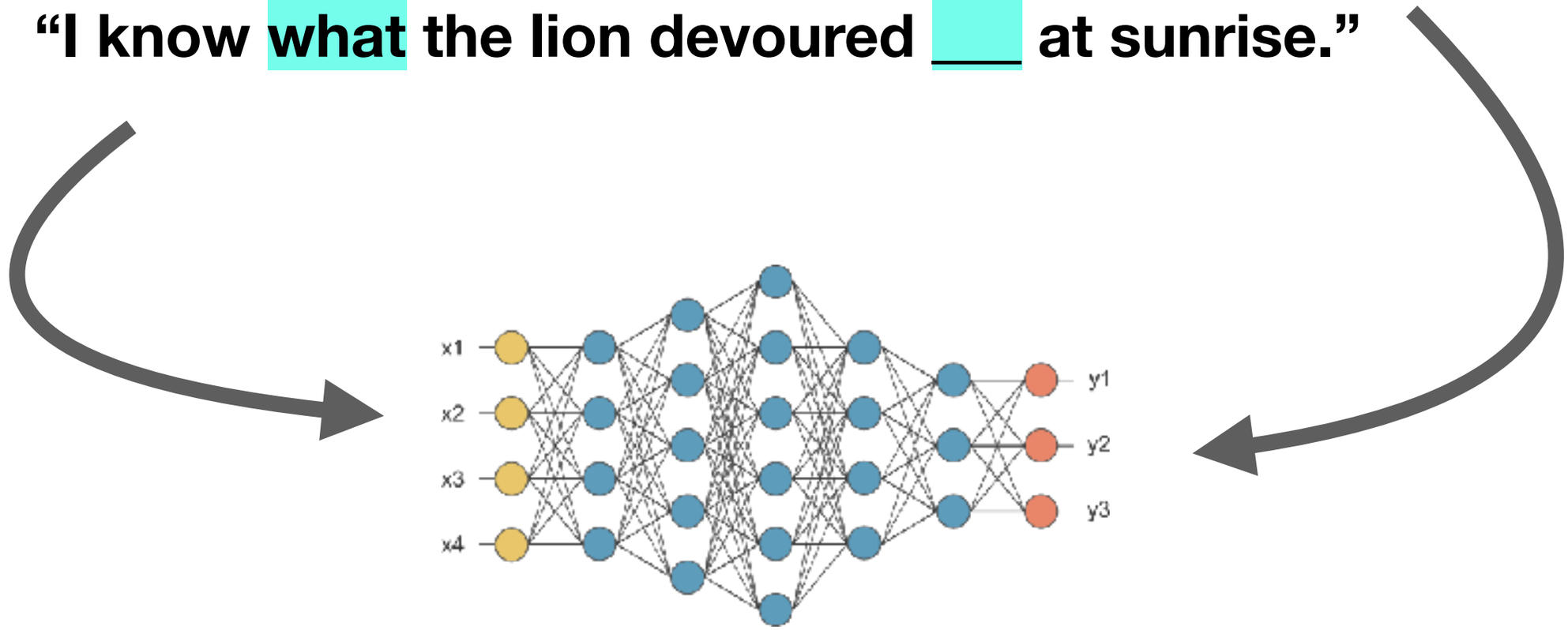



- Introduce the **wh-licensing interaction**

# Technical Motivation

## **Psycholinguistics Paradigm** for RNN Assessment

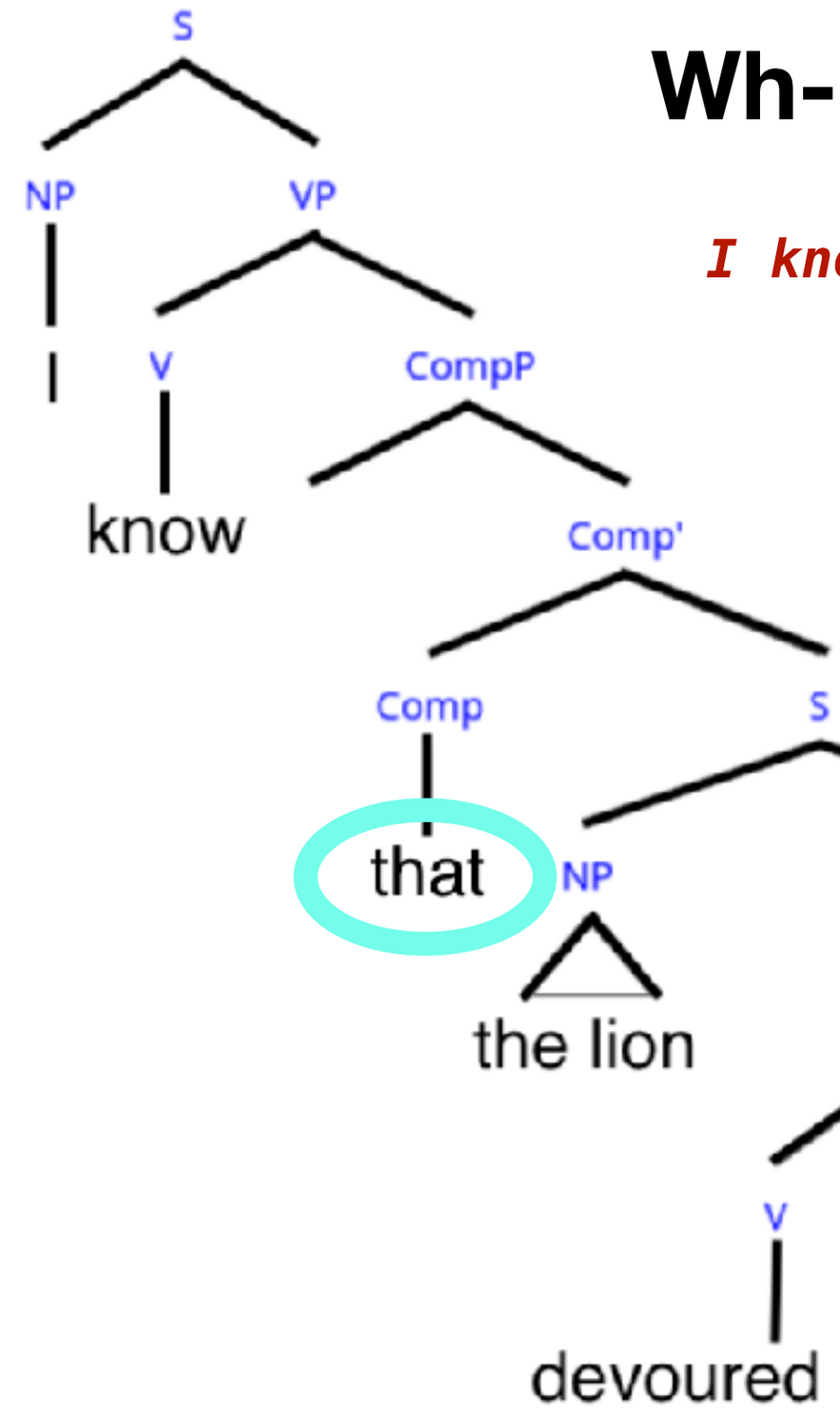
**“I know what the lion devoured ___ at sunrise.”**




- Introduce the **wh-licensing interaction**

- Do the models reason about **argument structure**?
- Do the models reason about **absence of material**?

# Methods: Wh-Licensing Interaction

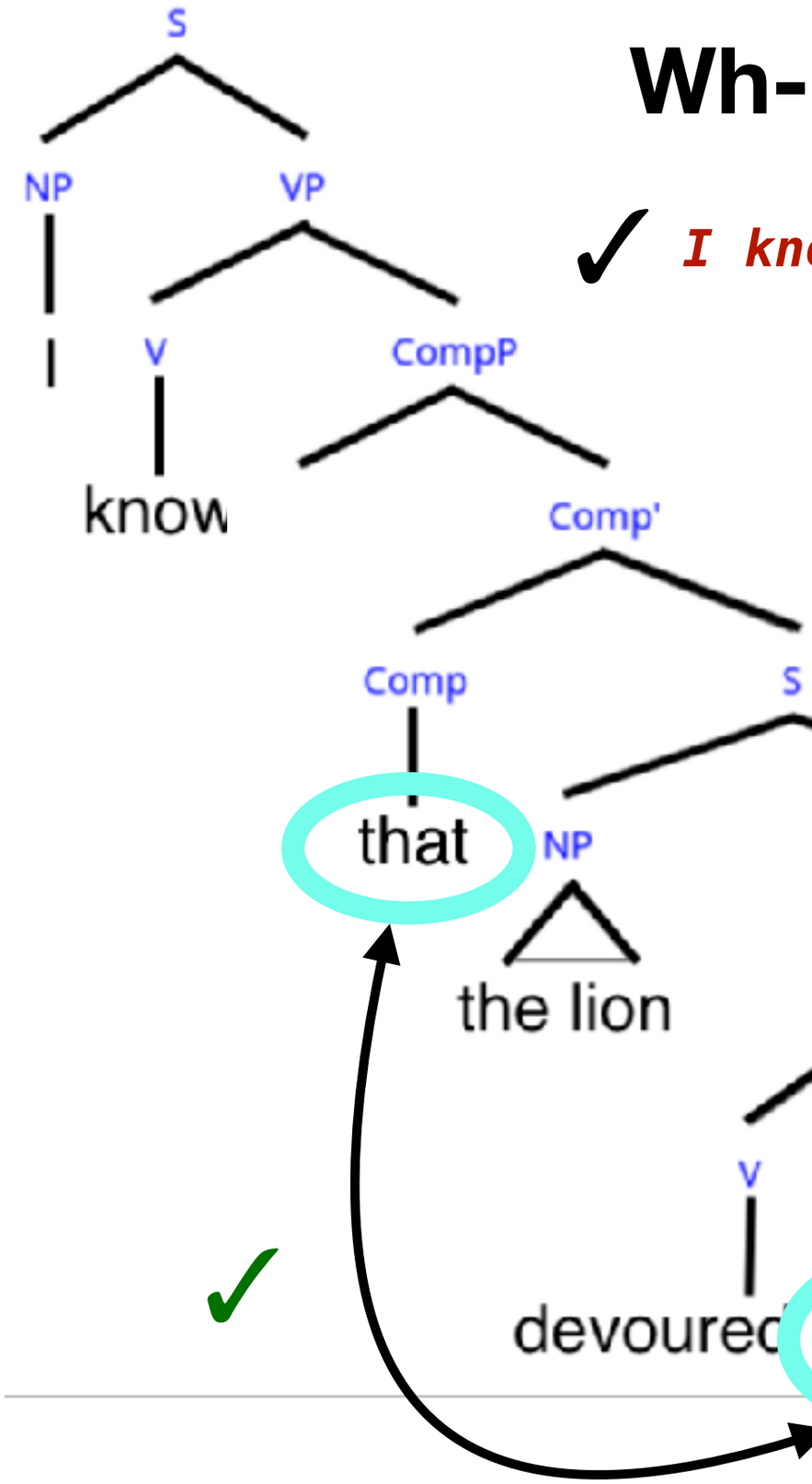



## Wh-Licensing Interaction

*I know that the lion devoured the gazelle at sunrise.*

**-FILLER**
**-GAP**

# Methods: Wh-Licensing Interaction

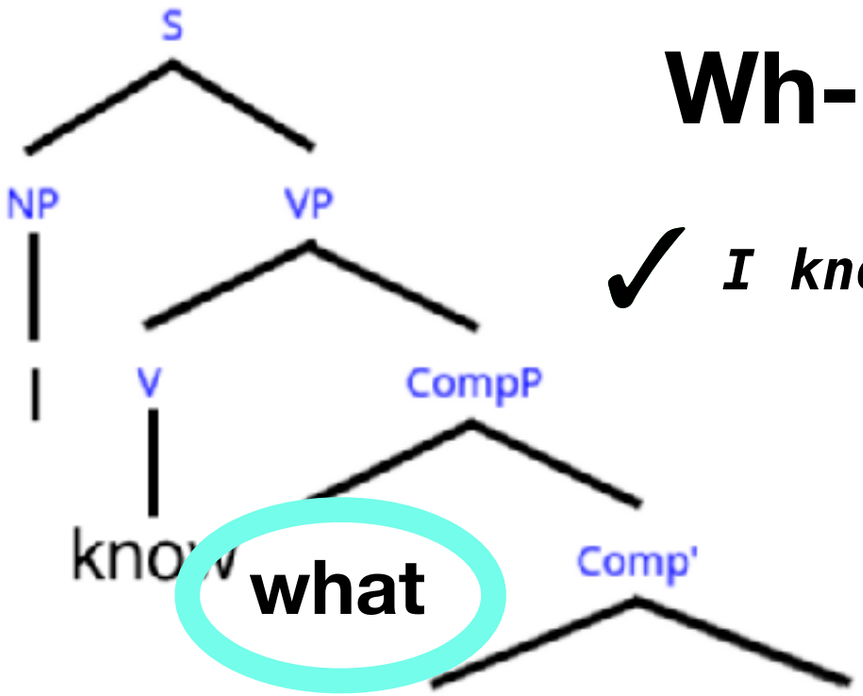



## Wh-Licensing Interaction

✓ *I know that the lion devoured the gazelle at sunrise.*

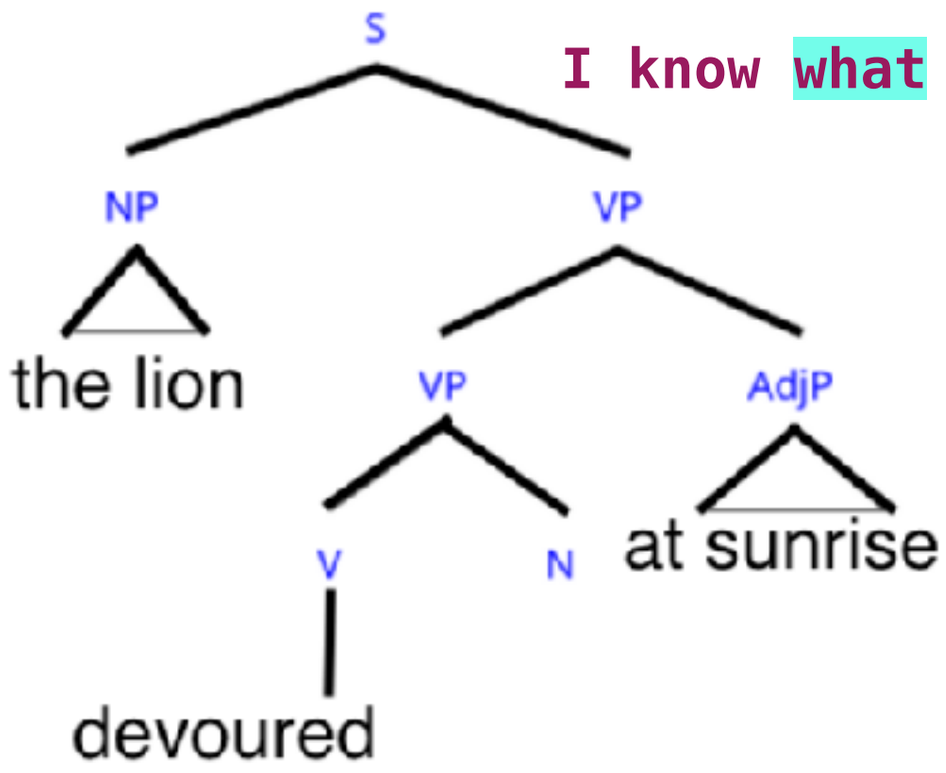
**-FILLER**
**-GAP**

# Methods: Wh-Licensing Interaction




## Wh-Licensing Interaction

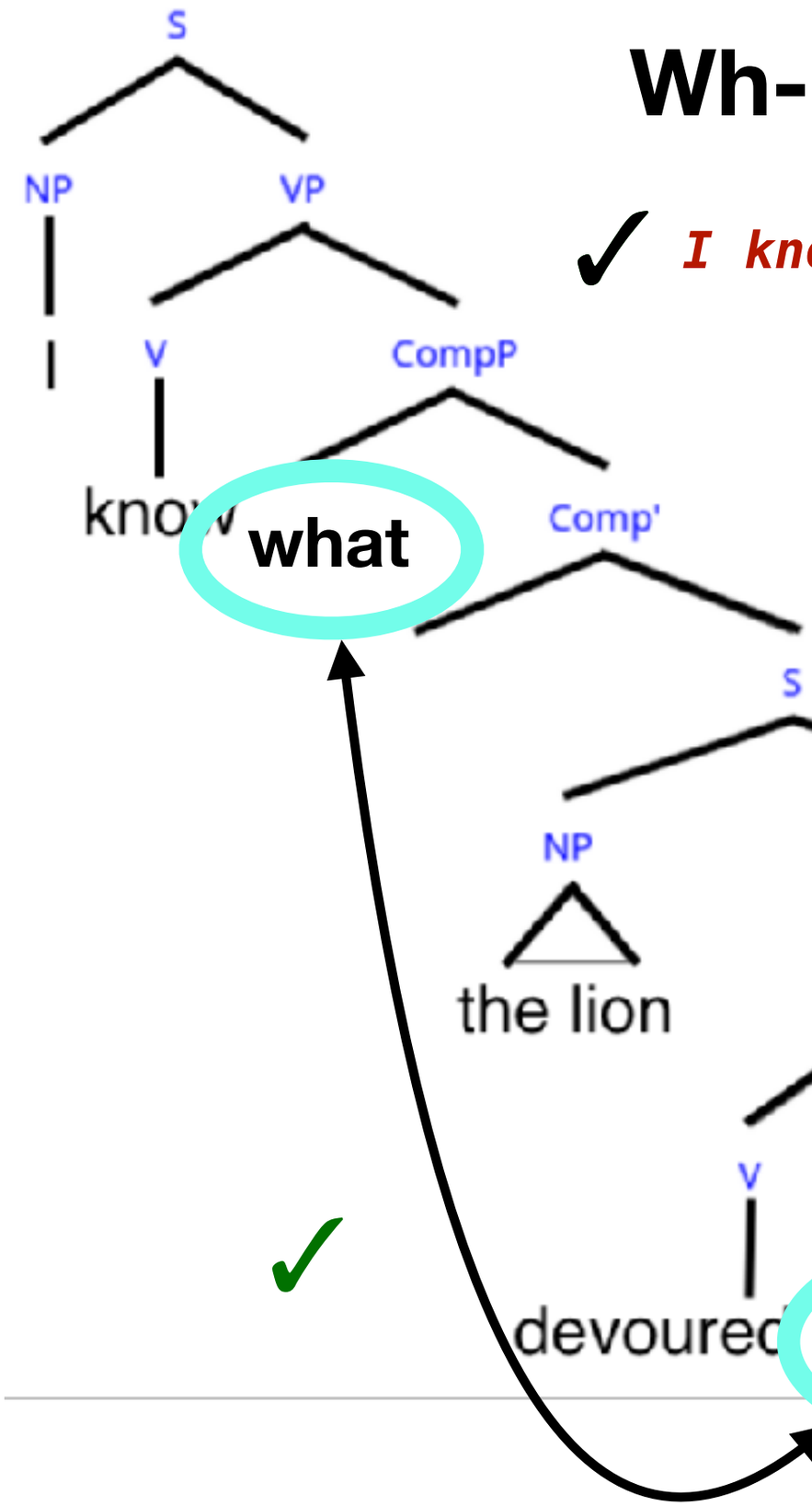
✓ *I know that the lion devoured the gazelle at sunrise.*




**I know what the lion devoured ___ at sunrise.**

**+FILLER**
**+GAP**

# Methods: Wh-Licensing Interaction



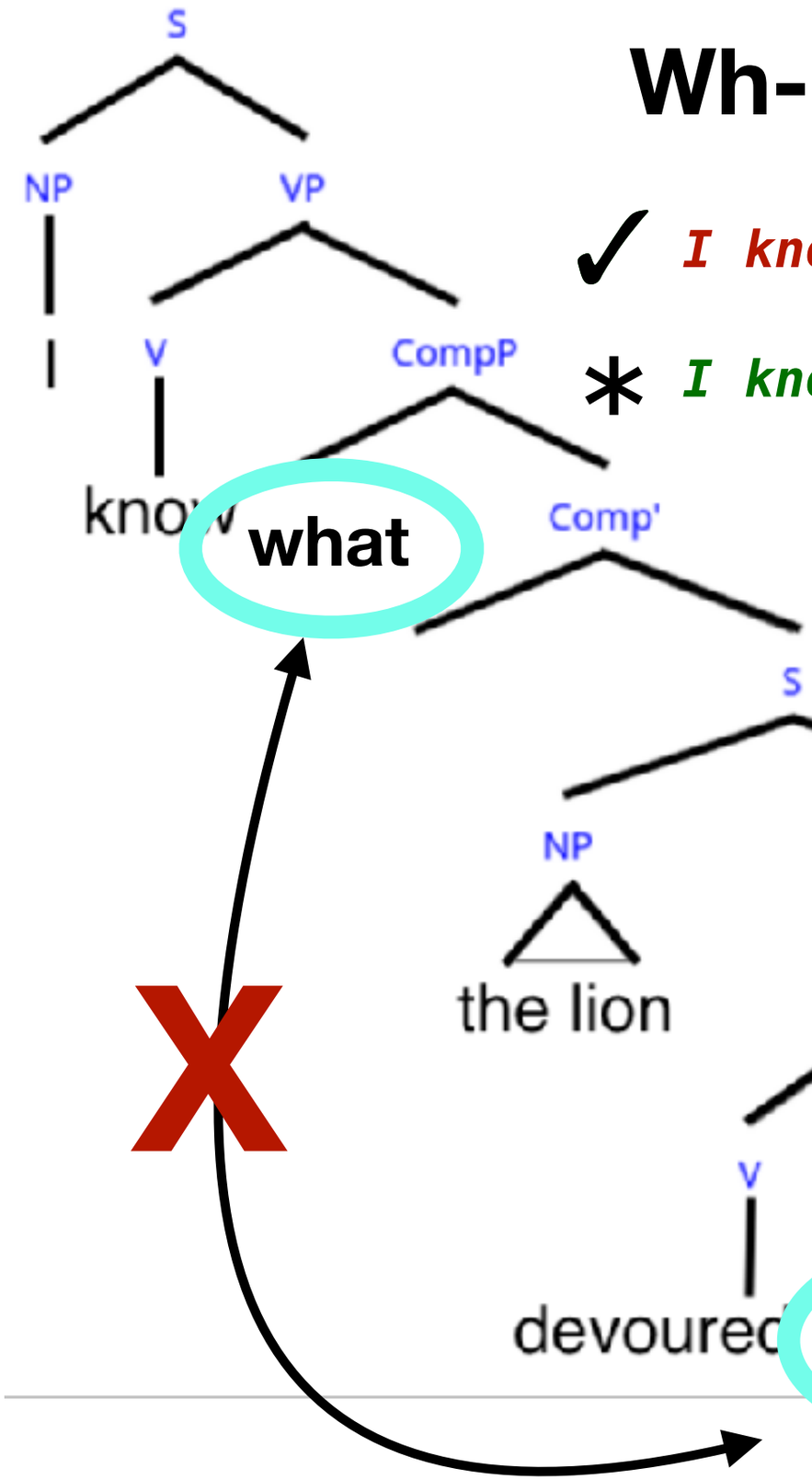

## Wh-Licensing Interaction

✓ *I know that the lion devoured the gazelle at sunrise.*

✓ I know what the lion devoured ___ at sunrise.

**+FILLER**
**+GAP**

# Methods: Wh-Licensing Interaction




## Wh-Licensing Interaction

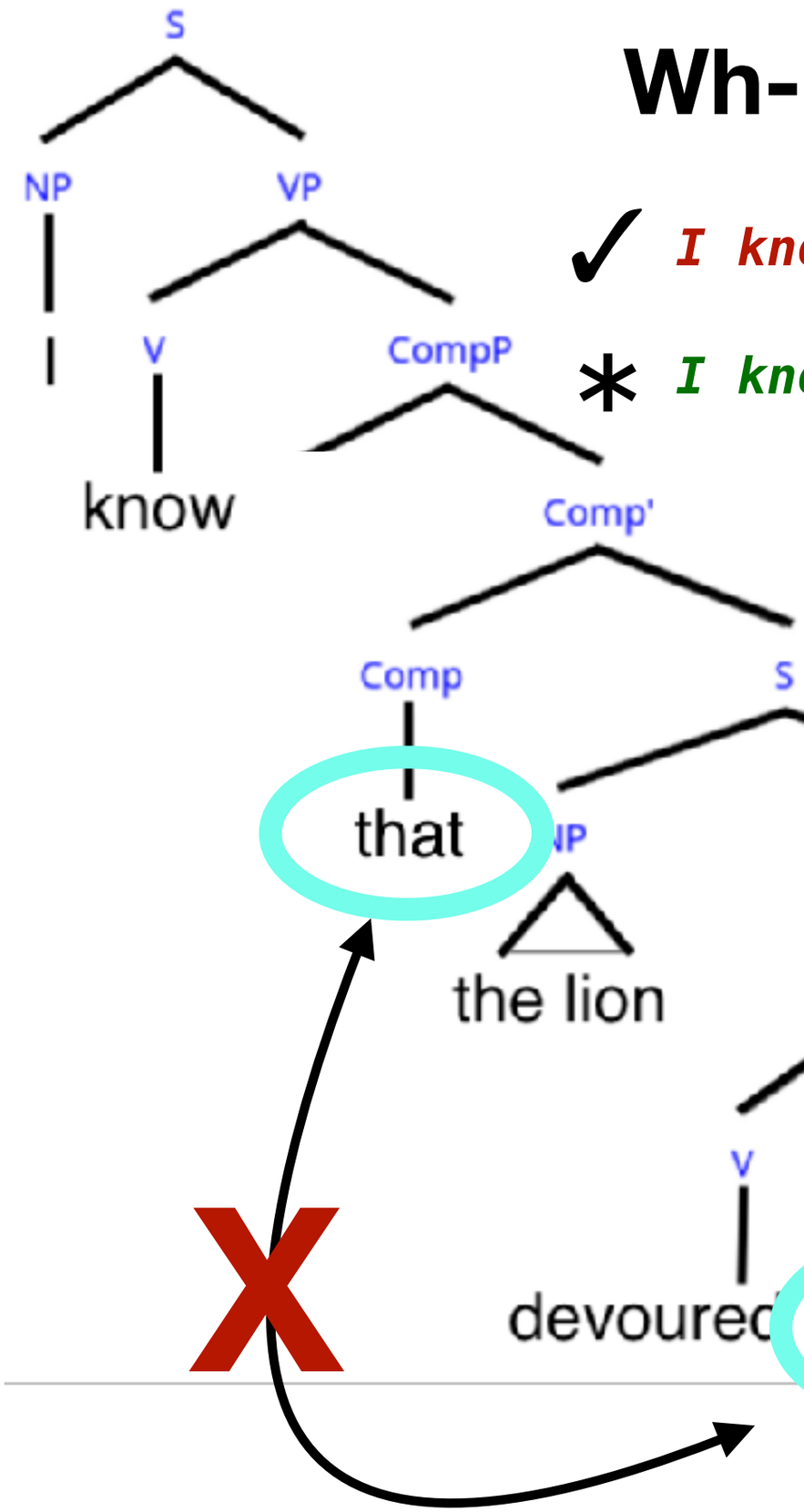
✓ *I know that the lion devoured the gazelle at sunrise.*

* *I know what the lion devoured the gazelle at sunrise.*

I know what the lion devoured ___ at sunrise.

**+FILLER**
**–GAP**

# Methods: Wh-Licensing Interaction




## Wh-Licensing Interaction

✓ *I know that the lion devoured the gazelle at sunrise.*

* *I know what the lion devoured the gazelle at sunrise.*

* I know that the lion devoured ___ at sunrise.

✓ I know what the lion devoured ___ at sunrise.

**–FILLER**
**+GAP**

# Methods: Wh-Licensing Interaction

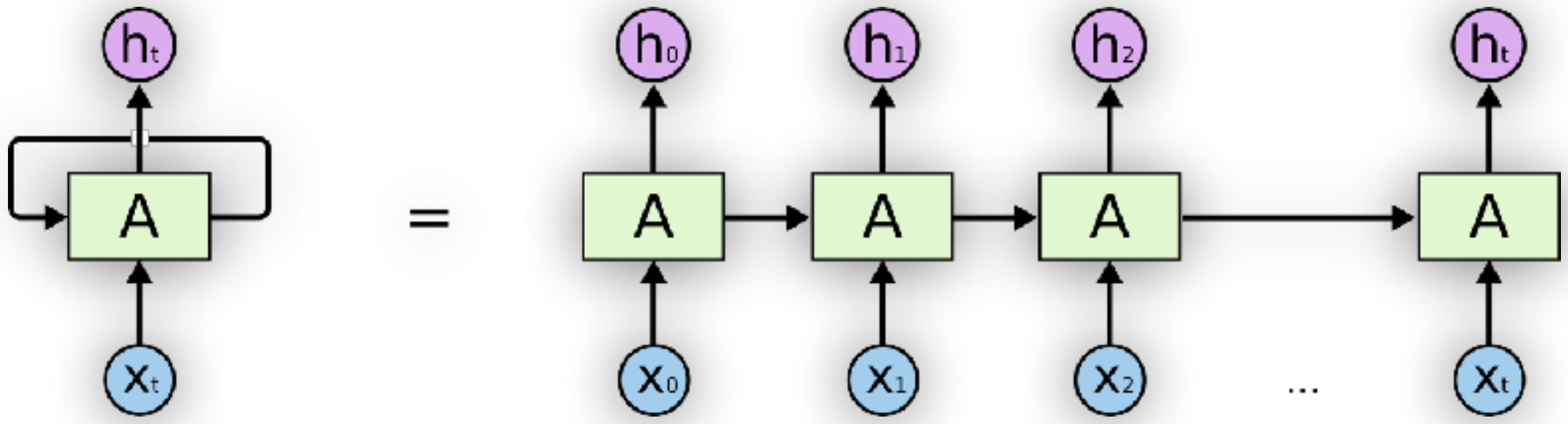
## Wh-Licensing Interaction

| (a) -FILLER -GAP | (b) +FILLER -GAP | (c) -FILLER +GAP | (d) +FILLER +GAP |
|---|---|---|---|

# Methods: Wh-Licensing Interaction

## Wh-Licensing Interaction

| (a) -FILLER -GAP | (b) +FILLER -GAP | (c) -FILLER +GAP | (d) +FILLER +GAP |
|---|---|---|---|
| $P(x_i \mid h_{i-1})$ | $P(x_i \mid h_{i-1})$ | $P(x_i \mid h_{i-1})$ | $P(x_i \mid h_{i-1})$ |



# Methods: Wh-Licensing Interaction

## Wh-Licensing Interaction

| (a) -FILLER -GAP | (b) +FILLER -GAP | (c) -FILLER +GAP | (d) +FILLER +GAP |
|---|---|---|---|
| $P(x_i \mid h_{i-1})$ | $P(x_i \mid h_{i-1})$ | $P(x_i \mid h_{i-1})$ | $P(x_i \mid h_{i-1})$ |
| *S(a)* | *S(b)* | *S(c)* | *S(d)* |

$$S(x_i) = -\log_2 p(x_i \mid h_{i-1})$$

# Methods: Wh-Licensing Interaction

## Wh-Licensing Interaction

| (a) -FILLER -GAP | (b) +FILLER -GAP | (c) -FILLER +GAP | (d) +FILLER +GAP |
|---|---|---|---|
| $P(x_i \mid h_{i-1})$ | $P(x_i \mid h_{i-1})$ | $P(x_i \mid h_{i-1})$ | $P(x_i \mid h_{i-1})$ |
| *S(a)* | *S(b)* | *S(c)* | *S(d)* |

S(b) - S(a)

**Fillers Set up Expectations for Gaps**

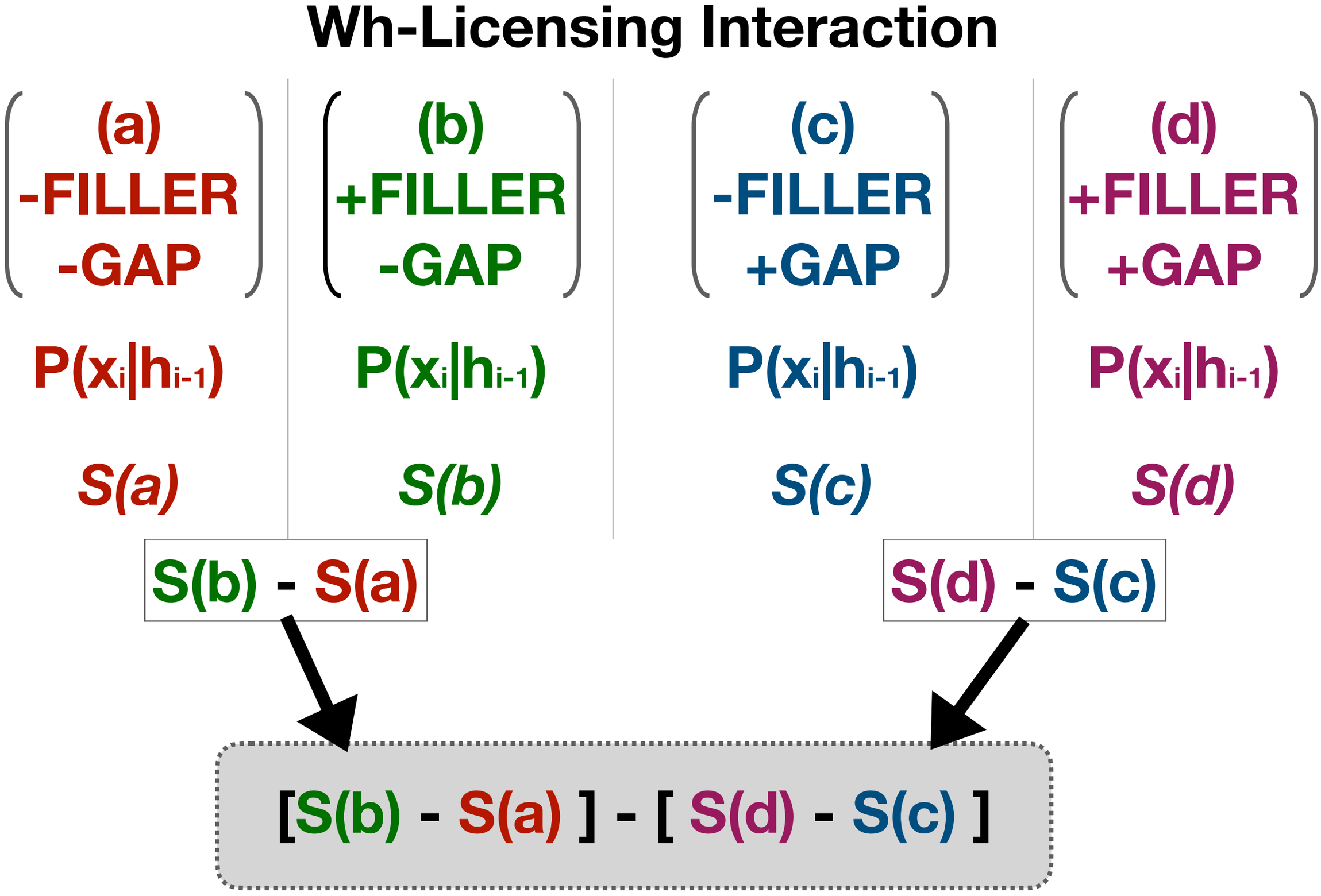
# Methods: Wh-Licensing Interaction

## Wh-Licensing Interaction

| (a) -FILLER -GAP | (b) +FILLER -GAP | (c) -FILLER +GAP | (d) +FILLER +GAP |
|---|---|---|---|
| $P(x_i \mid h_{i-1})$ | $P(x_i \mid h_{i-1})$ | $P(x_i \mid h_{i-1})$ | $P(x_i \mid h_{i-1})$ |
| *S(a)* | *S(b)* | *S(c)* | *S(d)* |
| | S(b) - S(a) | | S(d) - S(c) |

**Gaps Require Fillers**

# Methods: Wh-Licensing Interaction



# Methods: Wh-Licensing Interaction

## Wh-Licensing Interaction

| (a) -FILLER -GAP | (b) +FILLER -GAP | (c) -FILLER +GAP | (d) +FILLER +GAP |
|---|---|---|---|
| $P(x_i \mid h_{i-1})$ | $P(x_i \mid h_{i-1})$ | $P(x_i \mid h_{i-1})$ | $P(x_i \mid h_{i-1})$ |
| *S(a)* | *S(b)* | *S(c)* | *S(d)* |
| | S(b) - S(a) | S(d) - S(c) | |

**Wh-Licensing Interaction**

**[S(b) - S(a) ] - [ S(d) - S(c) ]**

# Methods: Models Used

| | ‘Google’ Model | ‘Gulordava’ Model |
|---|---|---|
| Layers | 2 | 2 |
| Units Per Layer | 8196 | 650 |
| Training Data | Billion Word Benchmark | 90-Million Tokens of English Wikipedia |

# Experiment 1: Wh-Licensing by Syntactic Position

*I know who, despite protocol…*

✓ *… ___ showed the slides to the guests yesterday.*
*[subject]*

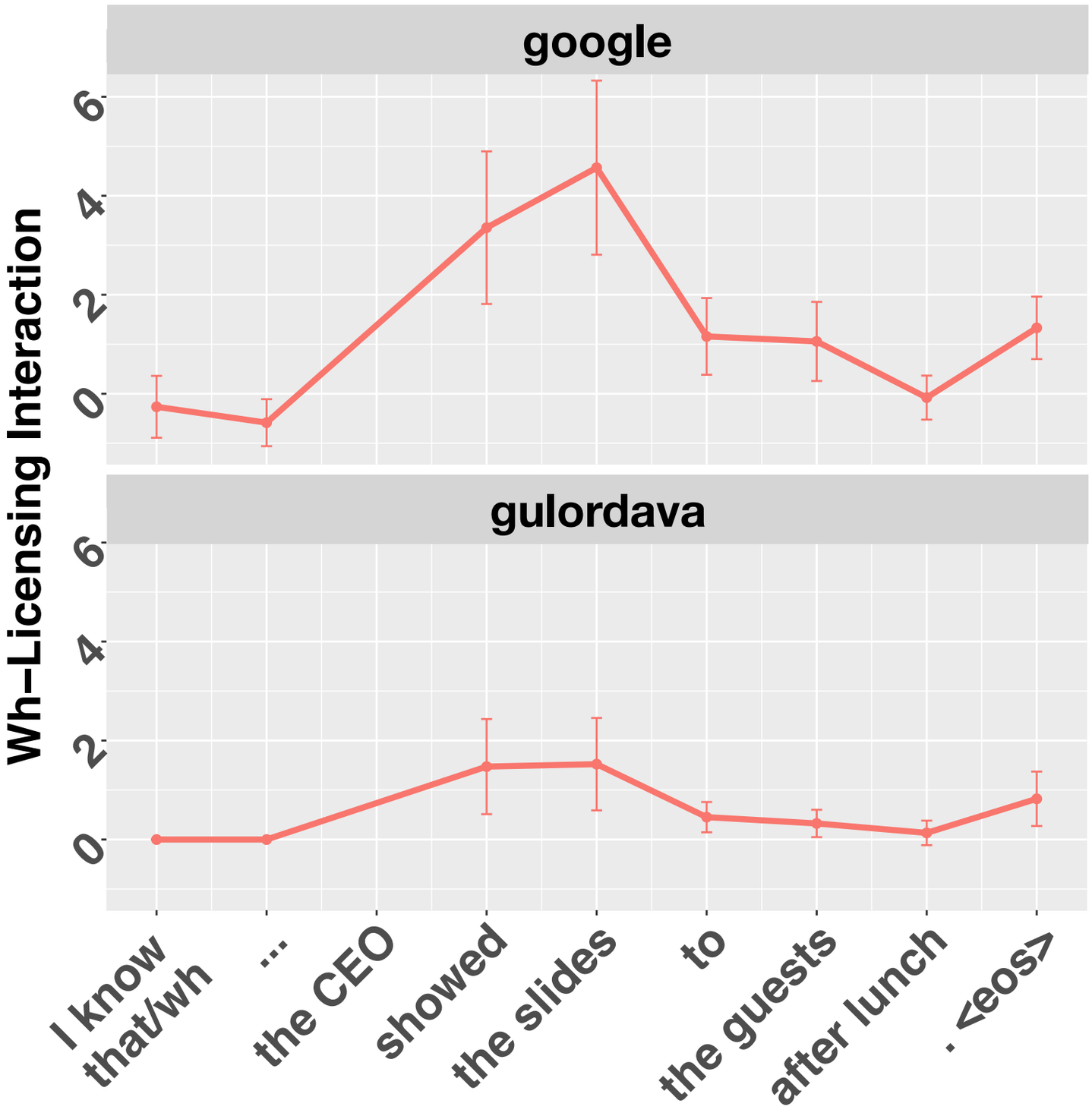
✓ *…the CEO showed ___ to the guests yesterday.*
*[object]*

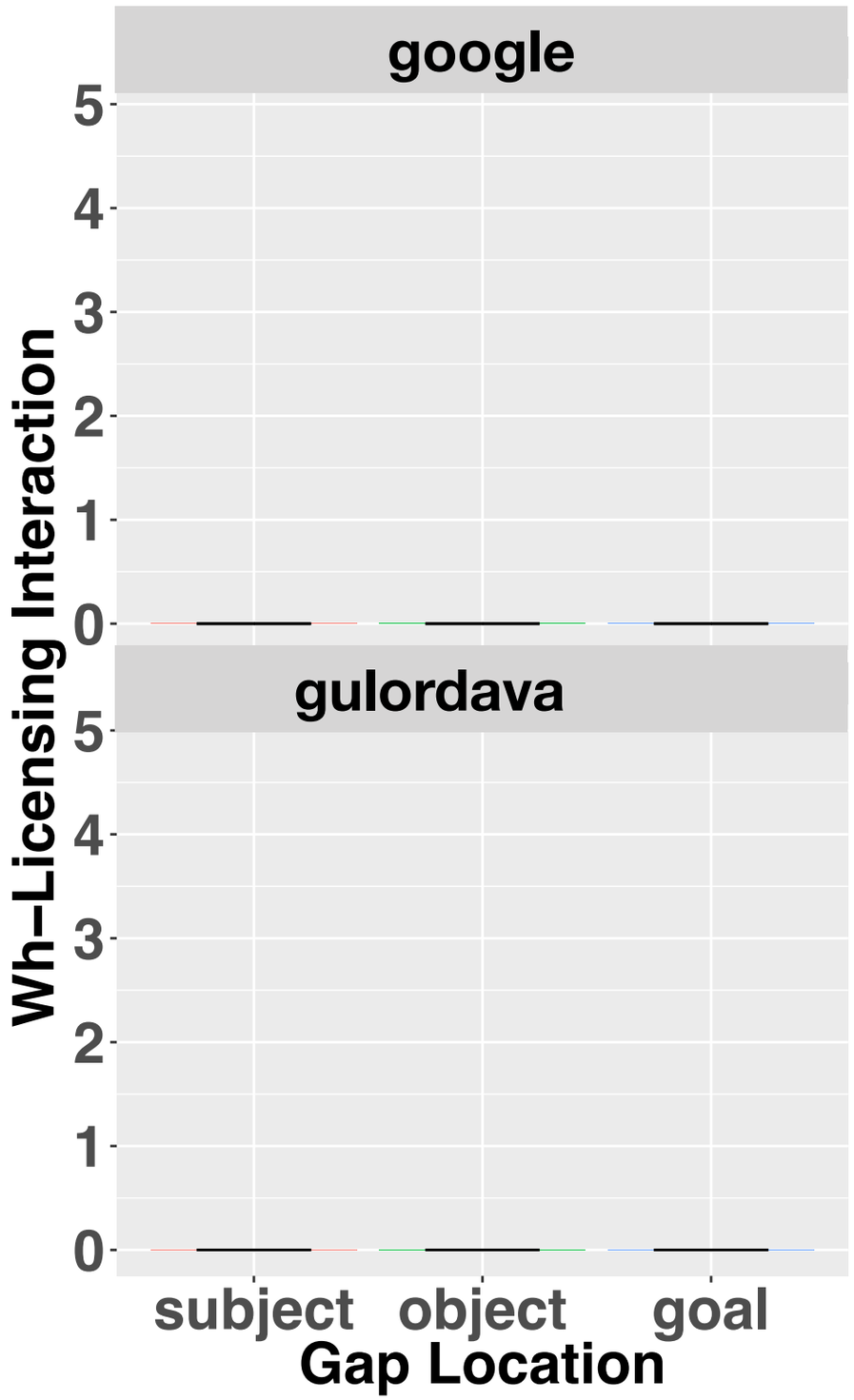
✓ *…the CEO showed the slides to ___ yesterday.*
*[goal]*

Do the RNNs Show Licensing Interaction in all 3 conditions?

# Experiment 1: Wh-Licensing by Syntactic Position

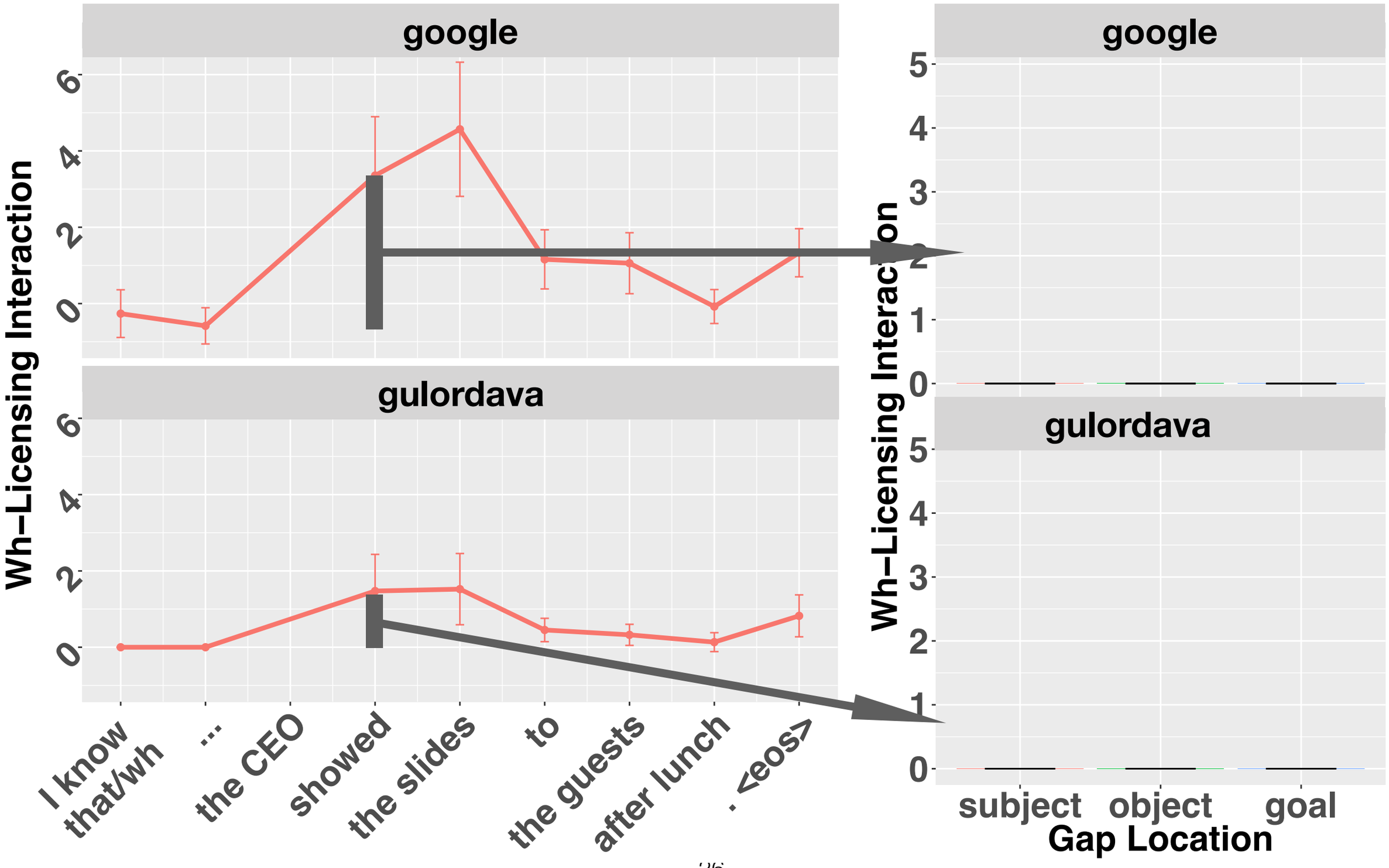
[subject]



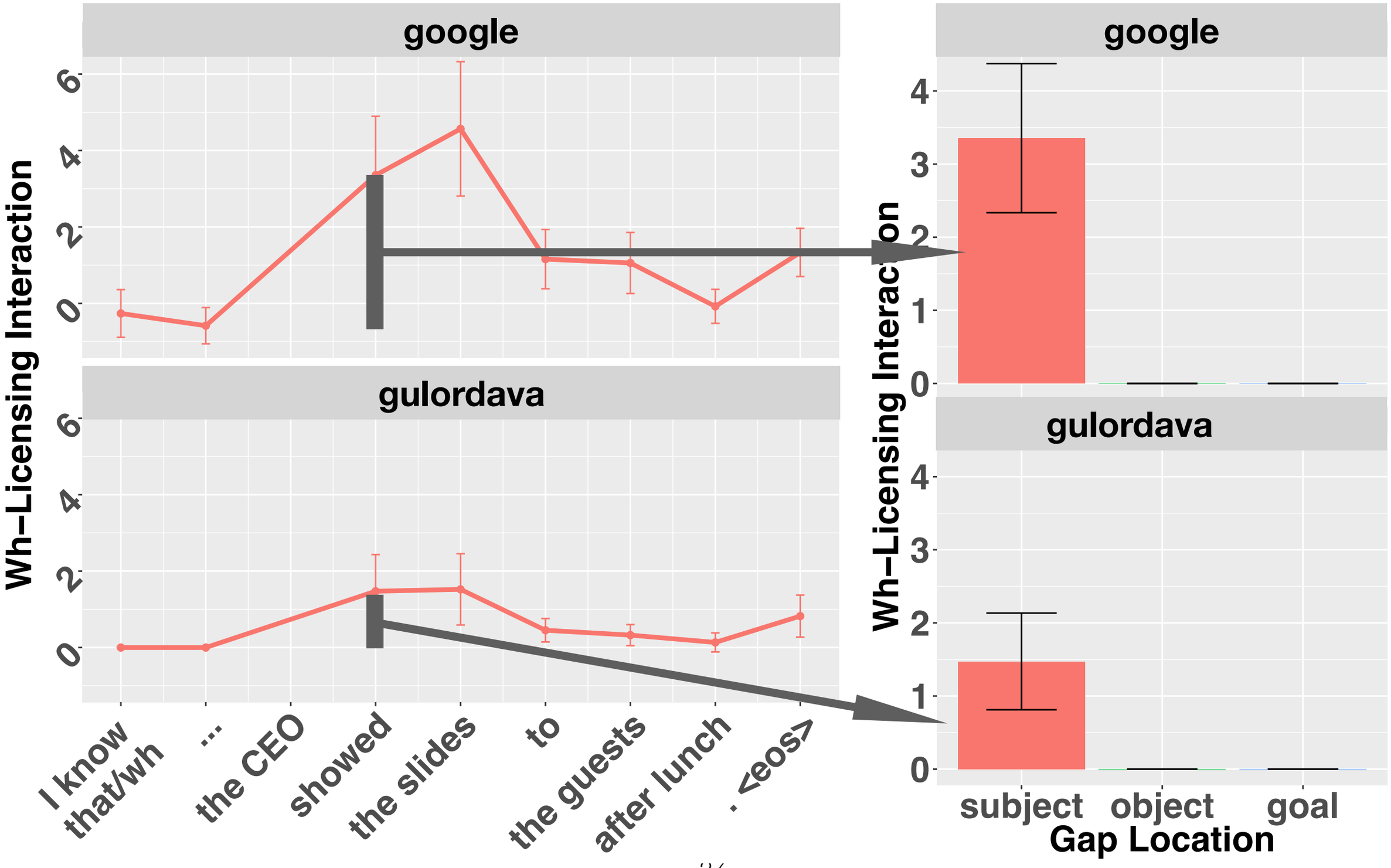
# Experiment 1: Wh-Licensing by Syntactic Position

**[subject]**



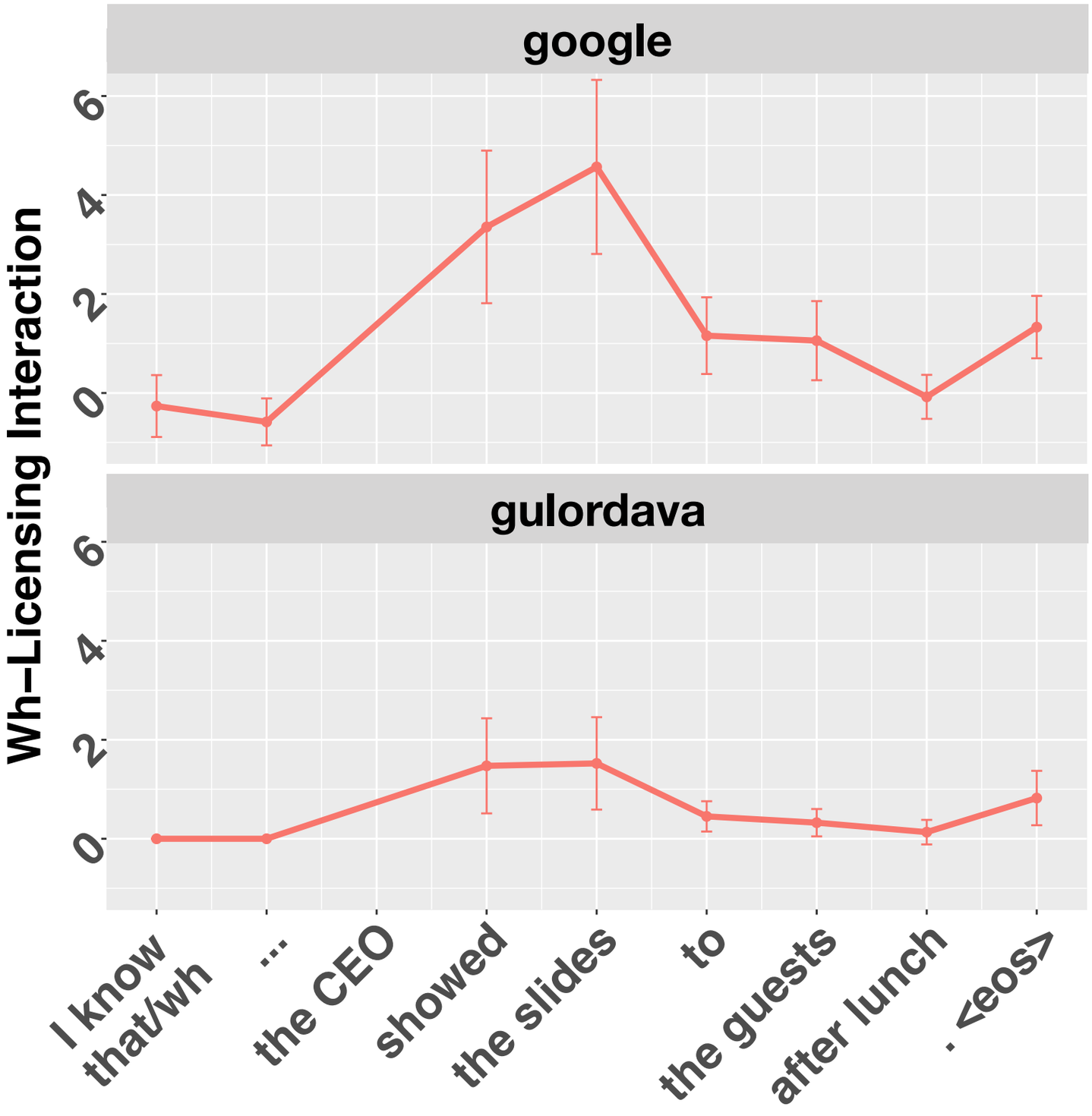
# Experiment 1: Wh-Licensing by Syntactic Position

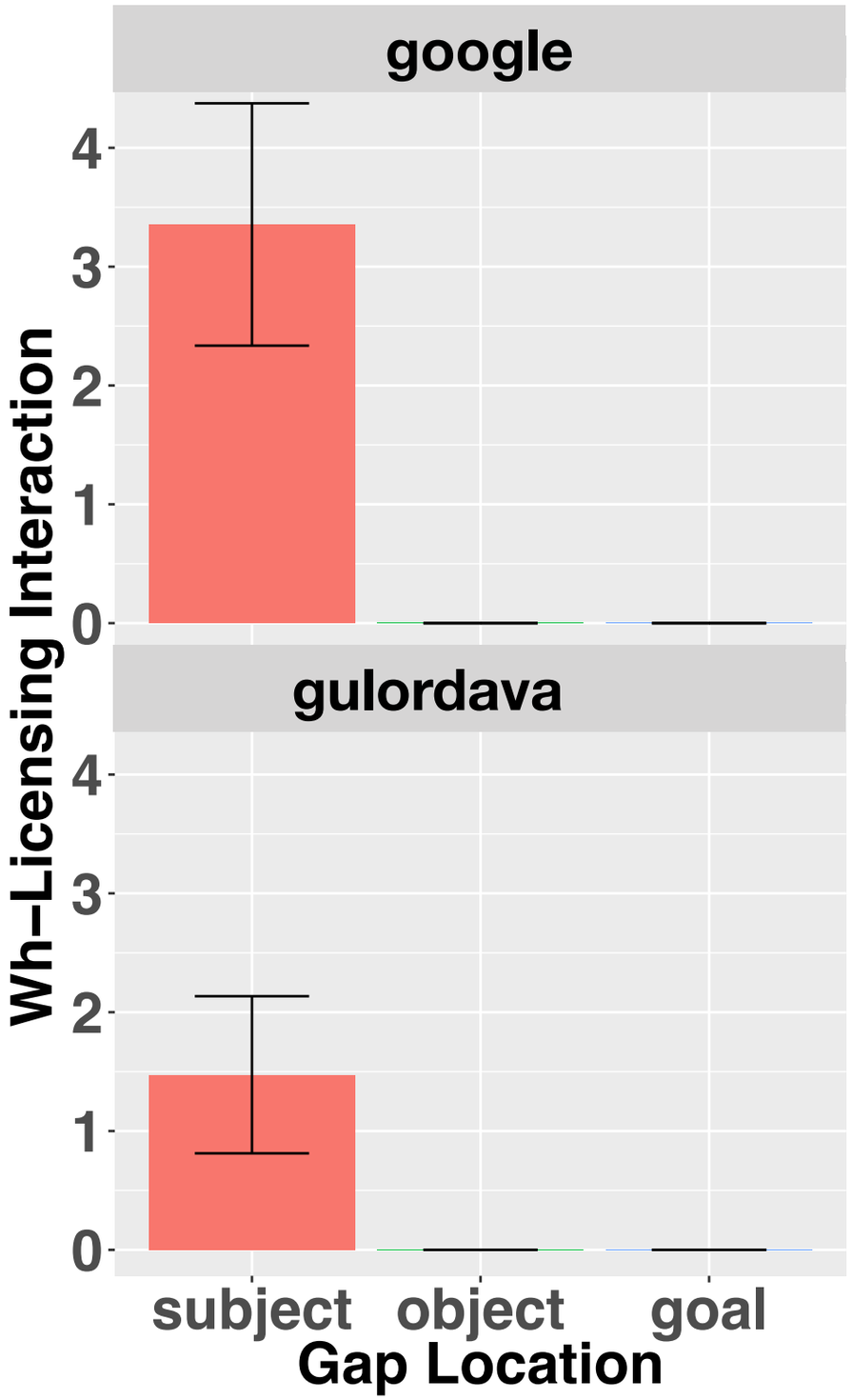
[subject]



# Experiment 1: Wh-Licensing by Syntactic Position

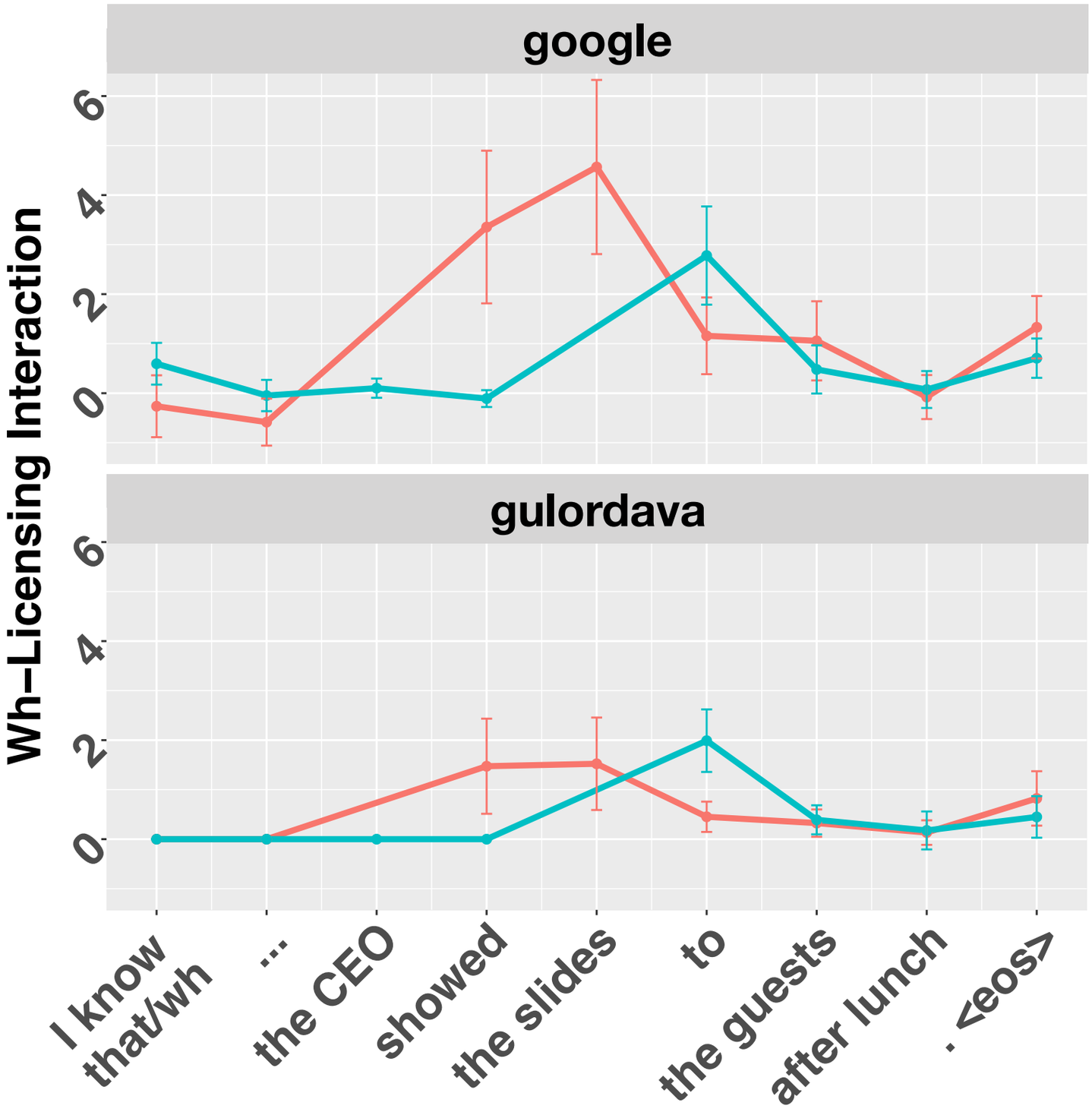
[subject]

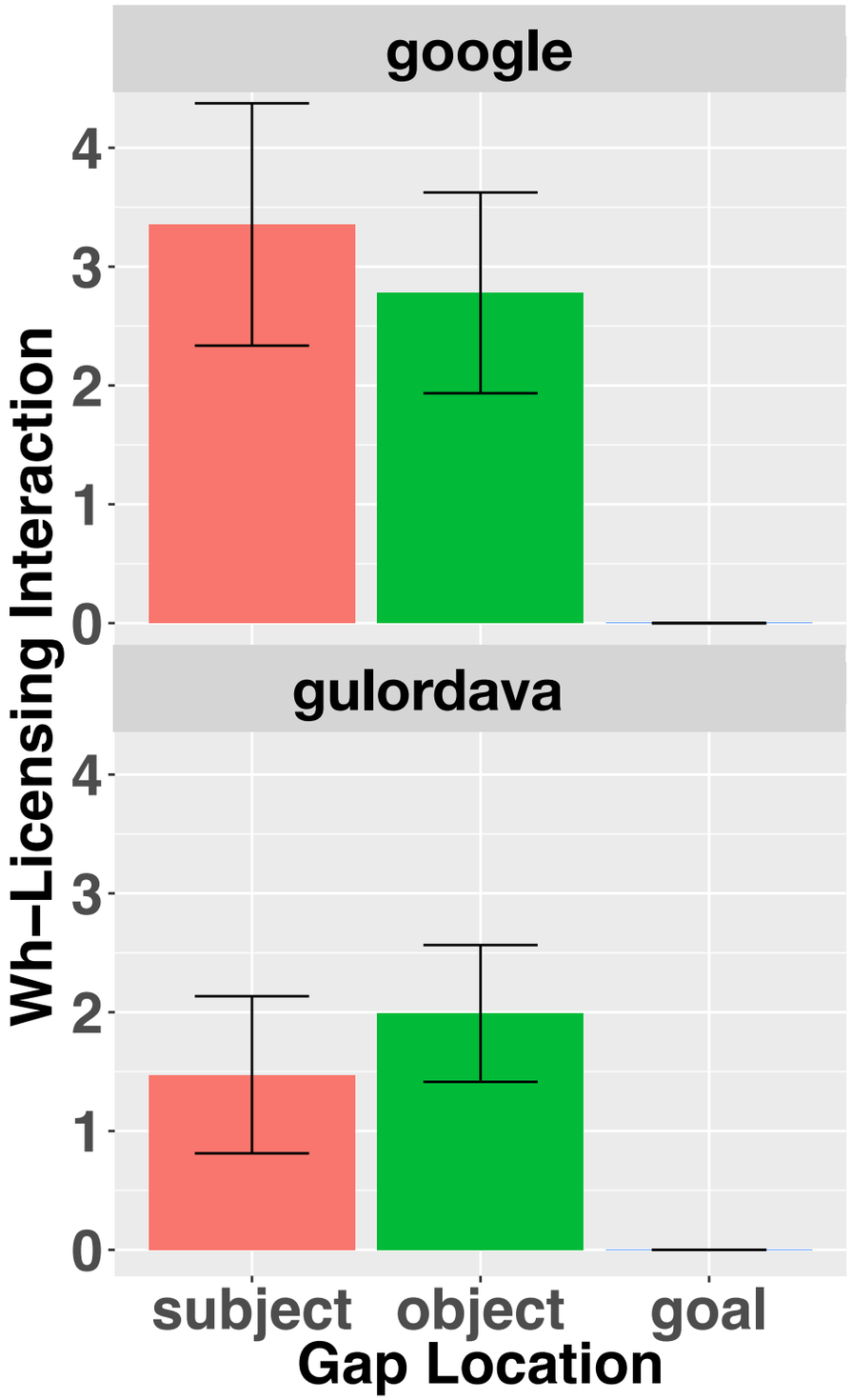


# Experiment 1: Wh-Licensing by Syntactic Position

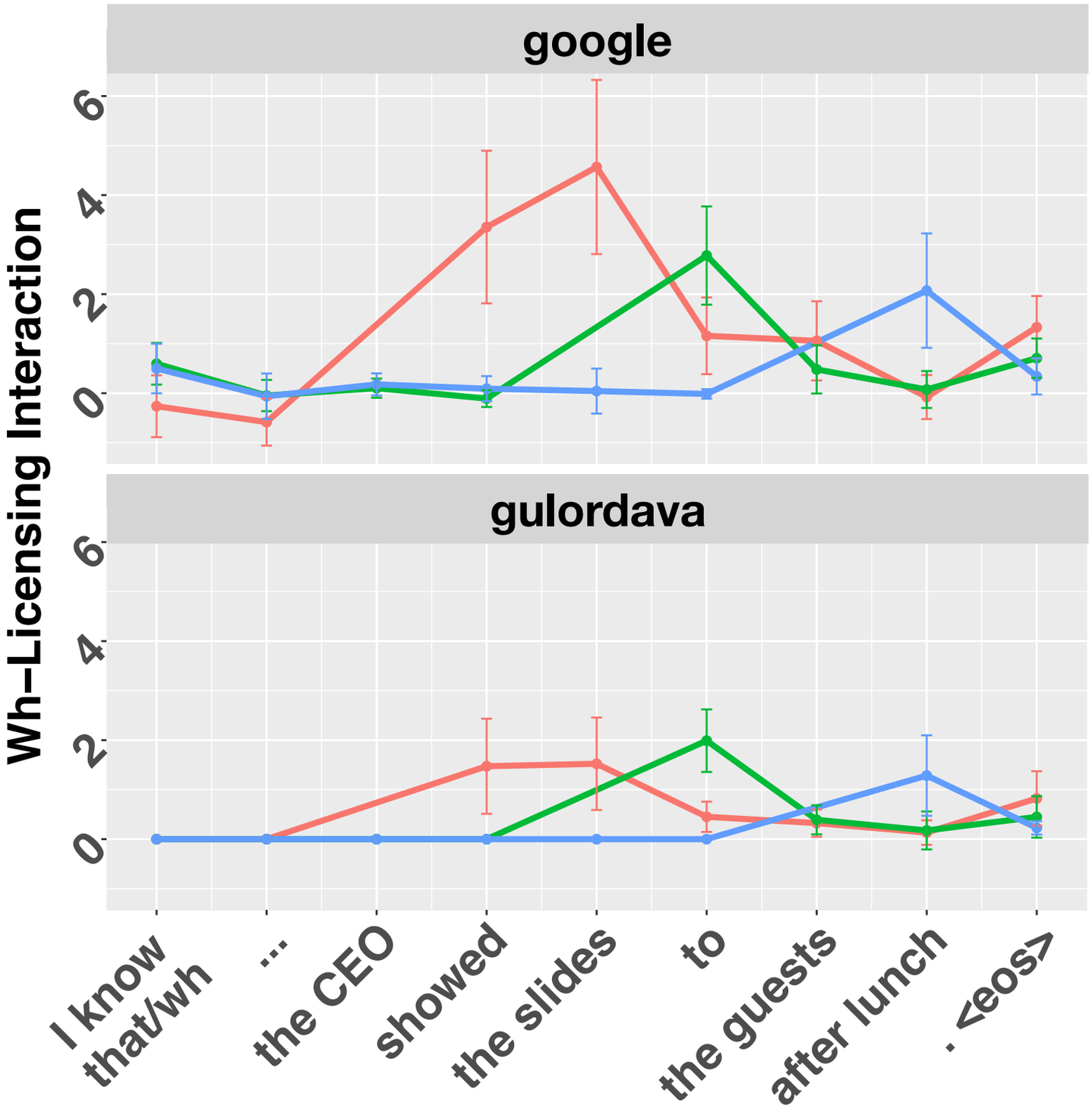
[object]

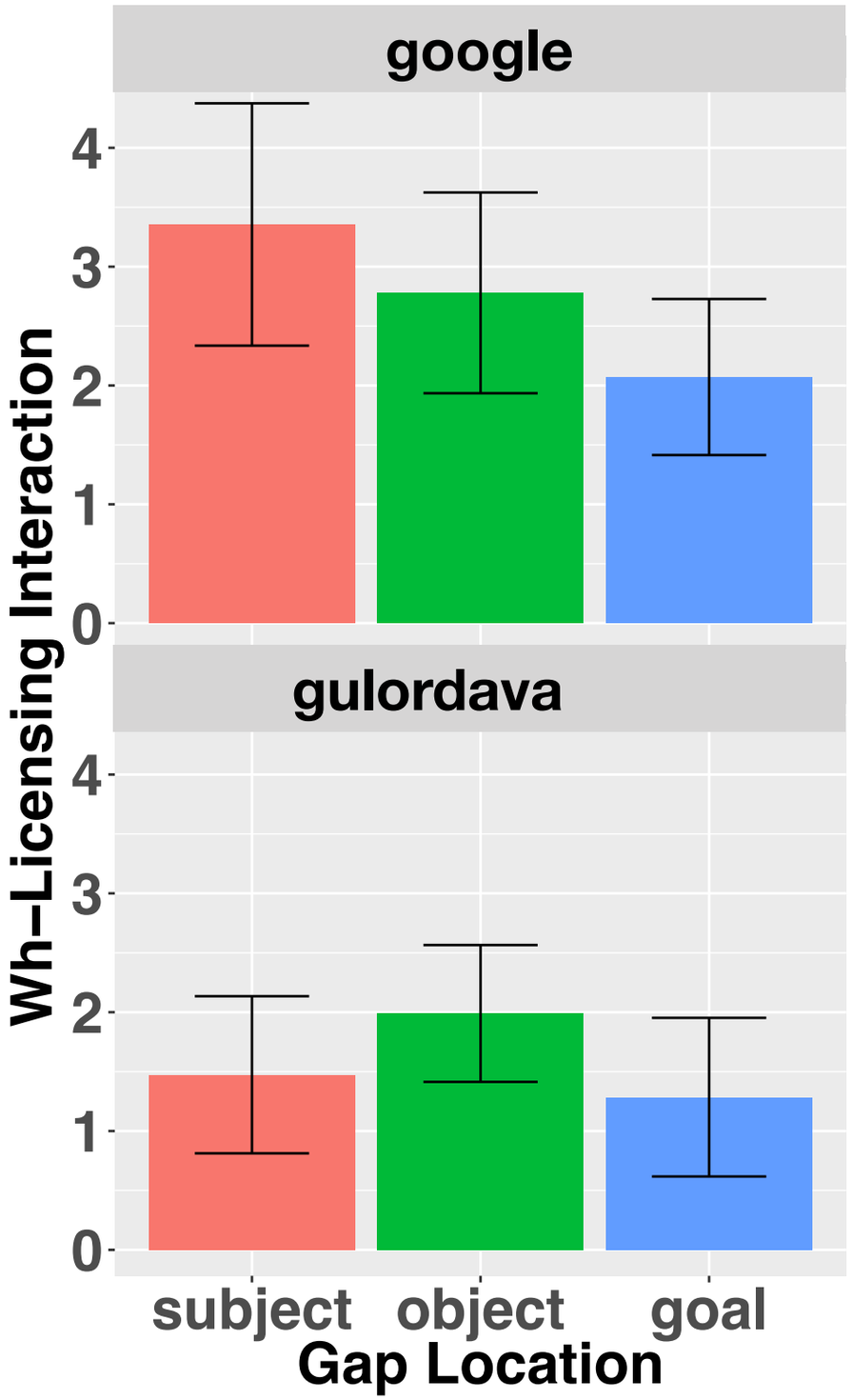


# Experiment 1: Wh-Licensing by Syntactic Position

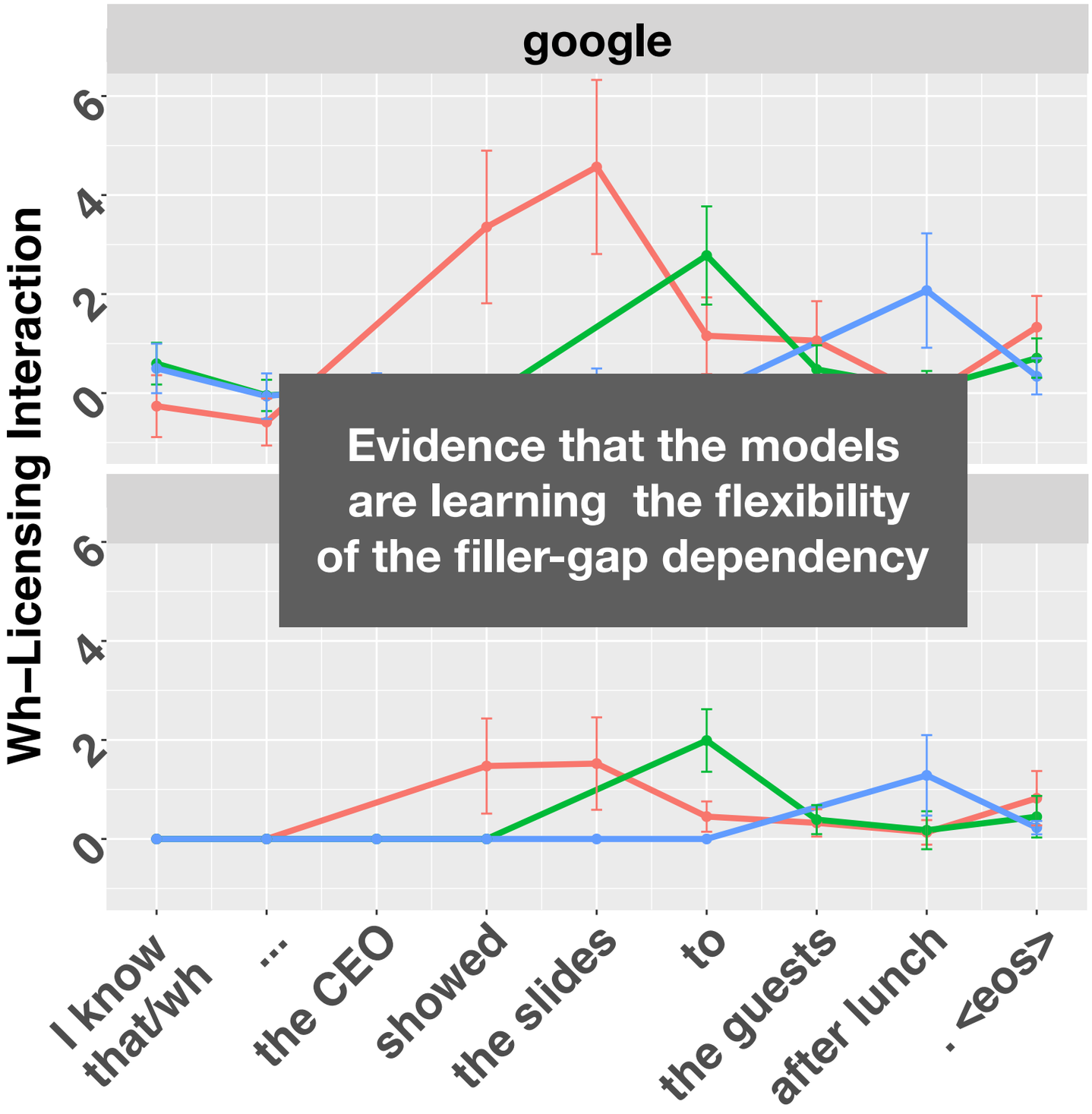
**[goal]**



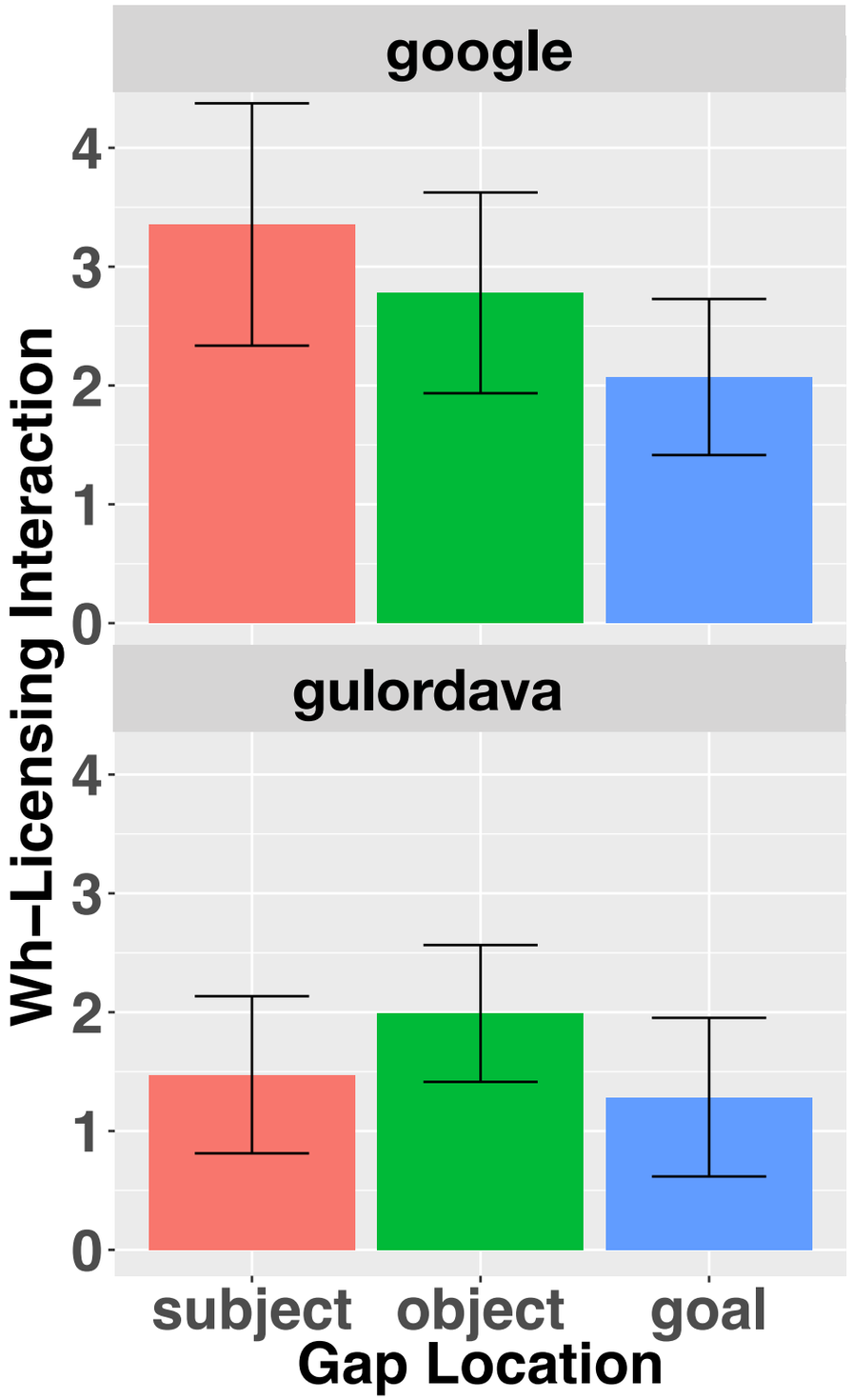



# Experiment 1: Wh-Licensing by Syntactic Position

[goal]




Evidence that the models are learning the flexibility of the filler-gap dependency



# Islands

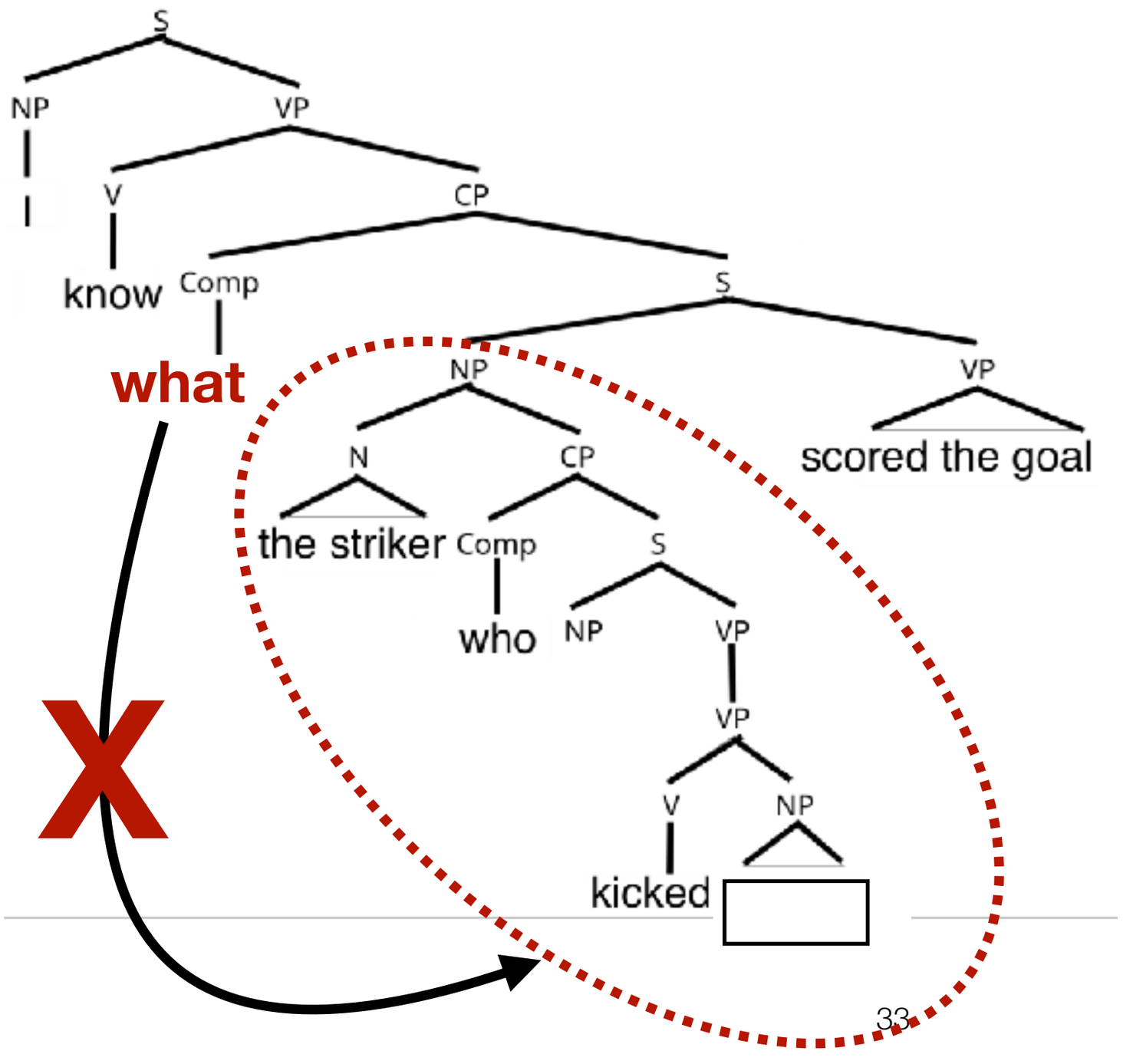
Syntactic arrangements that **block** the filler-gap dependency

✓ ***I know that the [striker who kicked the ball] scored the goal.***

# Islands

Syntactic arrangements that **block** the filler-gap dependency

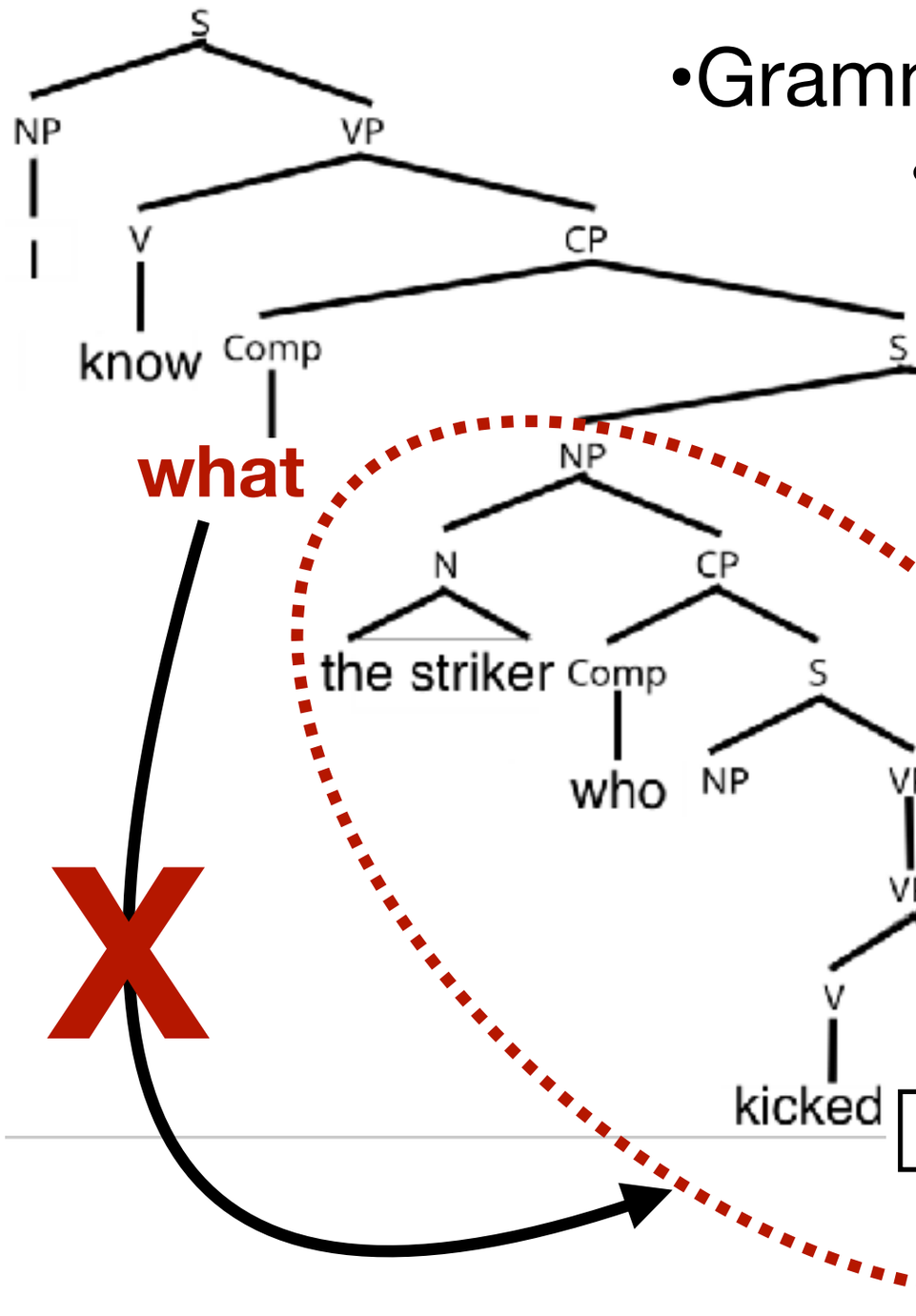
∗*I know* ***what*** *the [striker who kicked* __ *] scored the goal.*



# Islands

Syntactic arrangements that **block** the filler-gap dependency

∗*I know* **what** *the [striker who kicked __] scored the goal.*

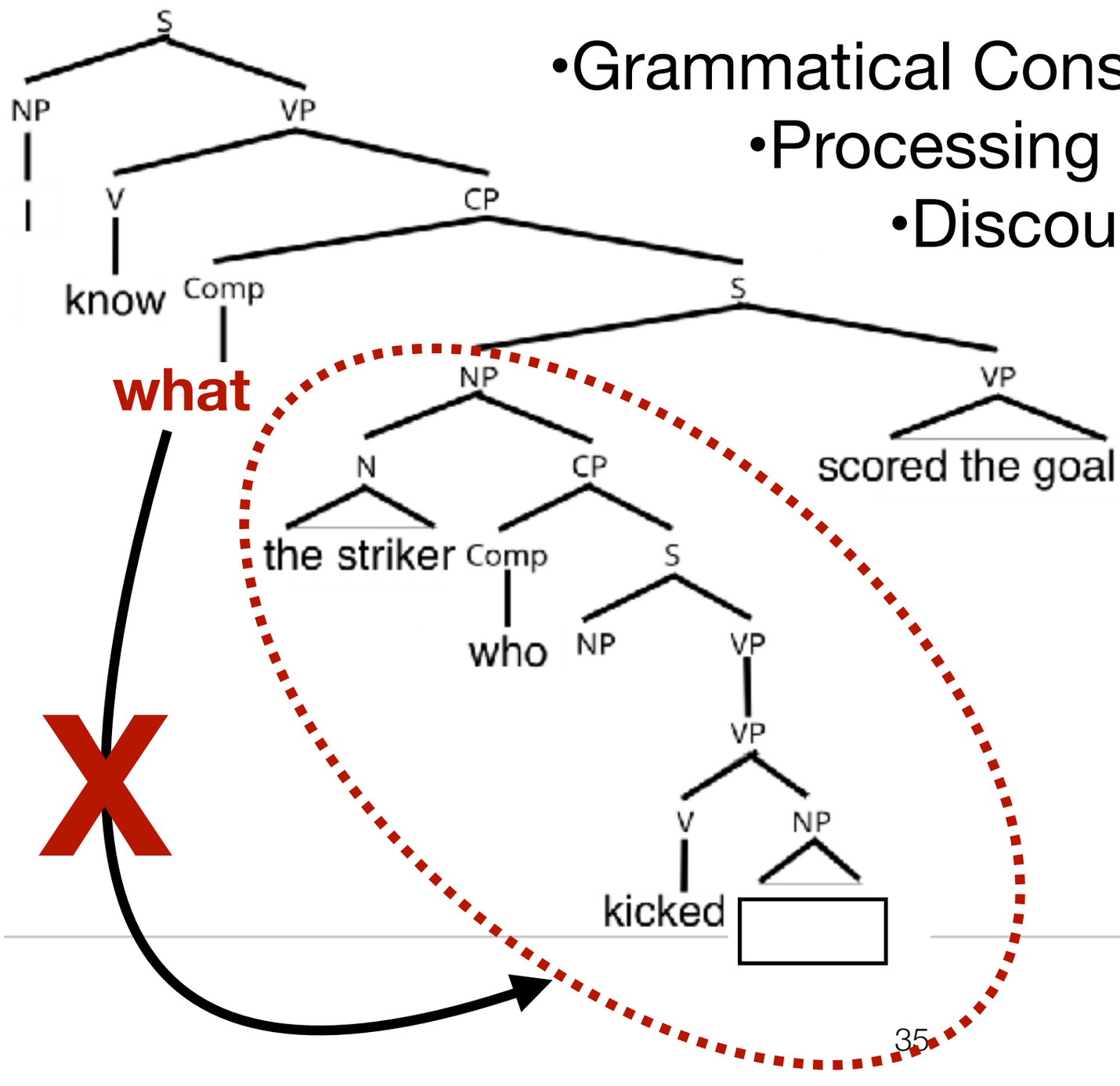



- Grammatical Constraint?
- Processing Difficulty?
- Discourse-Structural Factors?

# Islands

Syntactic arrangements that **block** the filler-gap dependency

∗*I know* ***what*** *the [striker who kicked* __ *] scored the goal.*




- Grammatical Constraint?
- Processing Difficulty?
- Discourse-Structural Factors?

**In this work:**

- **Adjunct Islands**
- **Wh-Islands**
- **Complex NP Islands**
- **Subject Islands**

# Experiment 2: Adjunct Islands

***Adjunct Clauses*** *block the filler—gap dependency*

*I know what…*

✓ *…the librarian in the glasses placed ___ on the wrong shelf.*

[object]

* *…the patron got mad **after** the librarian placed ___ on the wrong shelf.*

[*adjunct back*]

* ***…after** the librarian placed ___ on the wrong shelf, the patron got mad.*

[*adjunct front*]

Do the RNNs learn this constraint?

# Experiment 3: Wh-Islands

***Wh*-complementizers** block filler—gap dependencies:

*I know what Alex said…*

✓ *…your friend devoured ___ at the party.*
[null complementizer]

✓ ***…that** your friend devoured ___ at the party.*
[*that* complementizer]

* ***…whether** your friend devoured ___ at the party.*
[*wh*-complementizer]

Do the RNNs learn this?

# Experiment 4: Complex NP Islands

***Relative Clauses modifying Nouns*** *block the Filler—Gap dependency*

*I know who…*

✓*…my friend saw* ___ *at the museum.*

[object]

**…my friend saw a painting* ***that*** *depicted* ___ *at the museum.*

[*that cNP*]

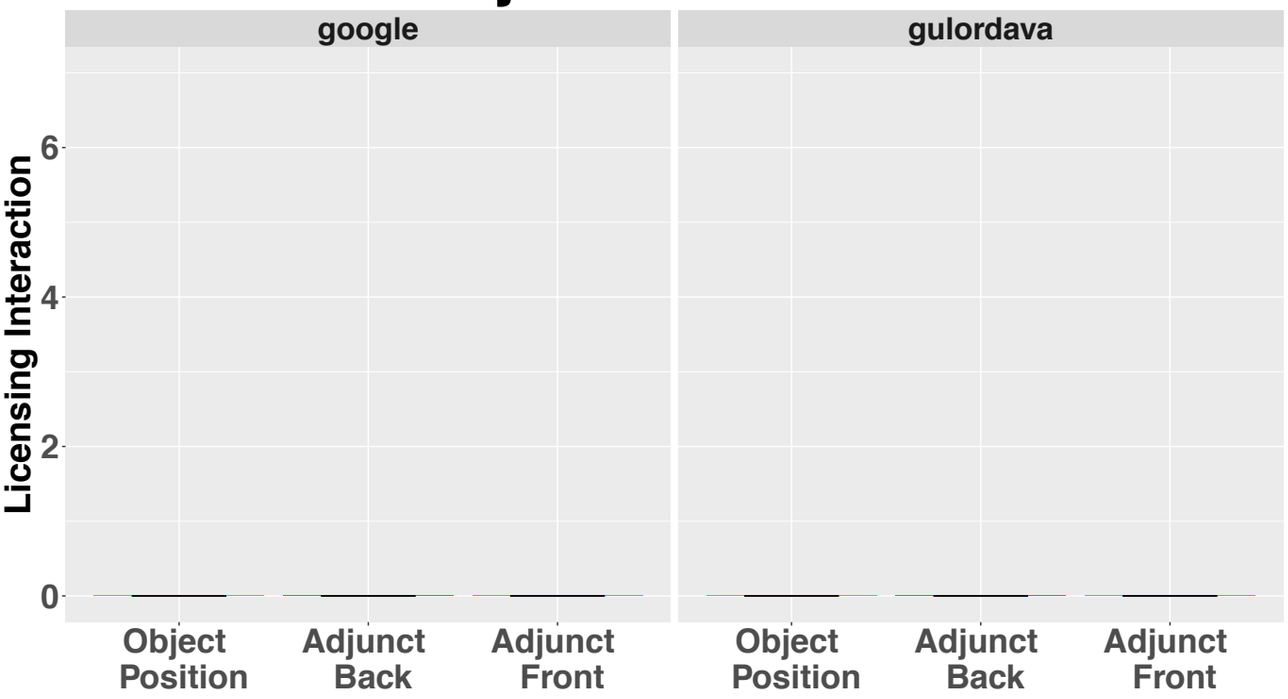
**…my friend saw a painting* ***which*** *depicted* ___ *at the museum.*

[*wh* cNP]

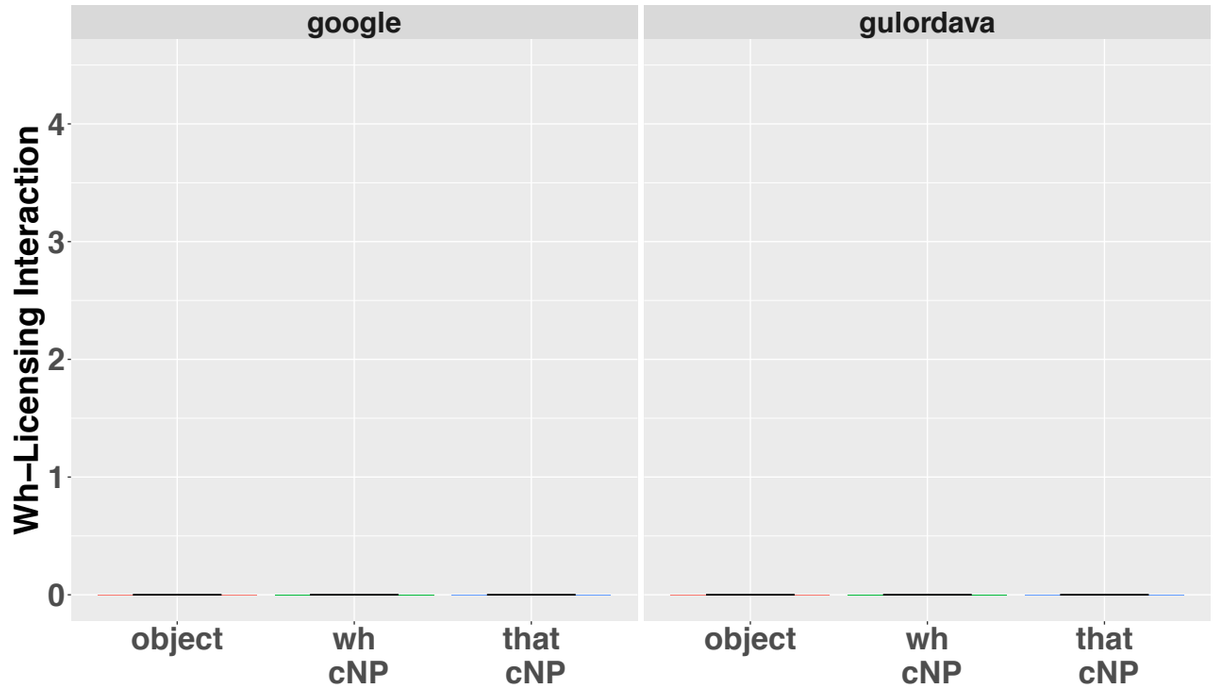
Do the RNNs learn this?

# Experiment 2-4: Results

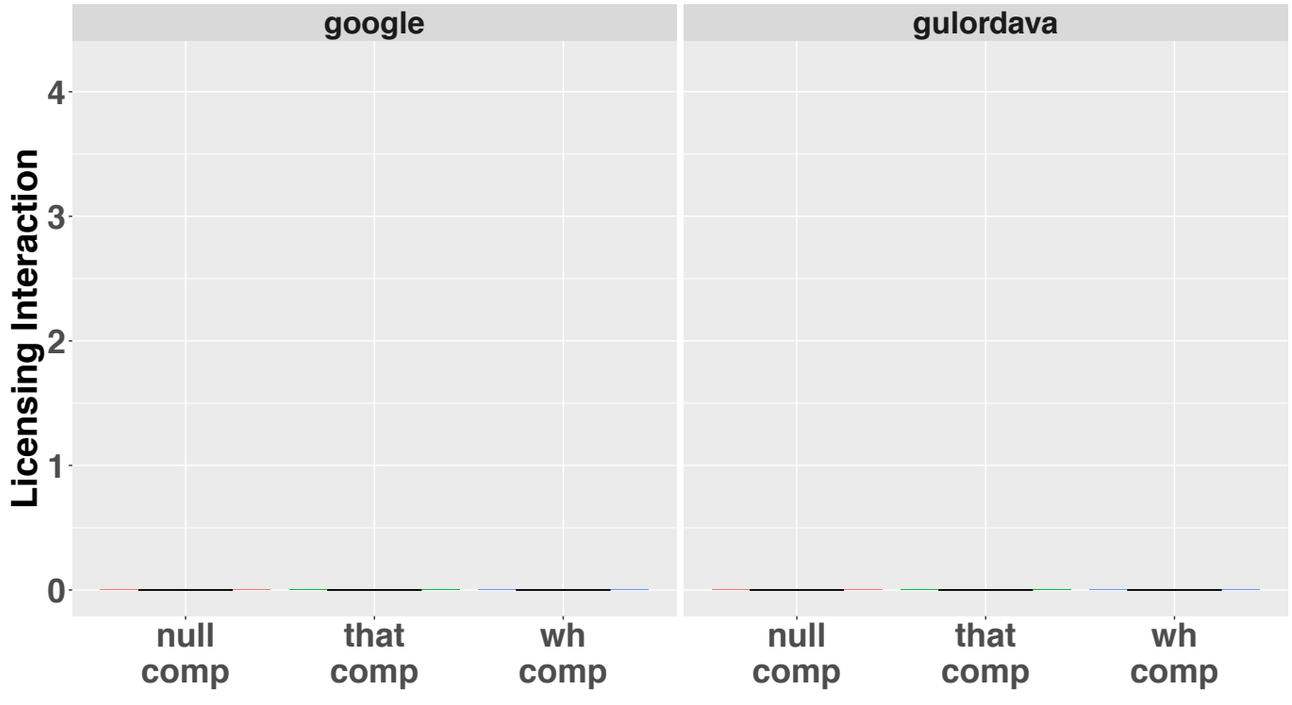
## Adjunct Islands



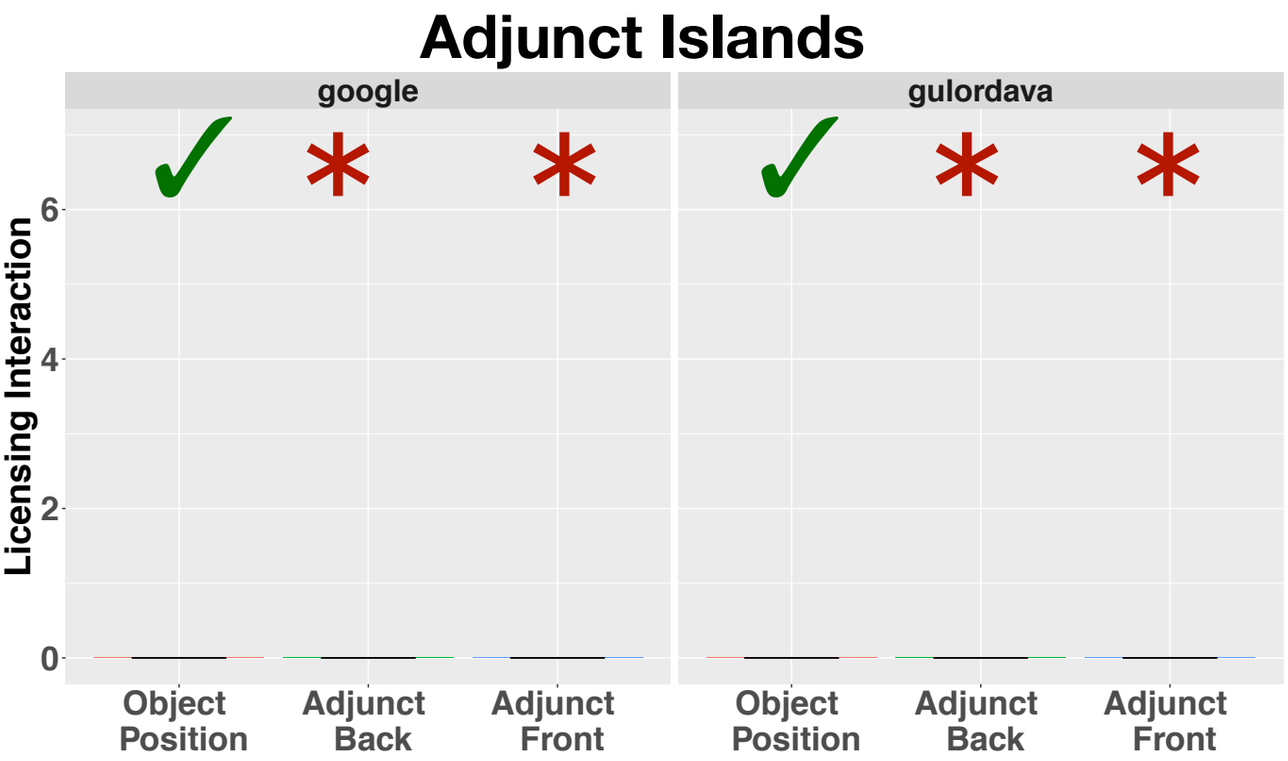

## Complex NP Islands

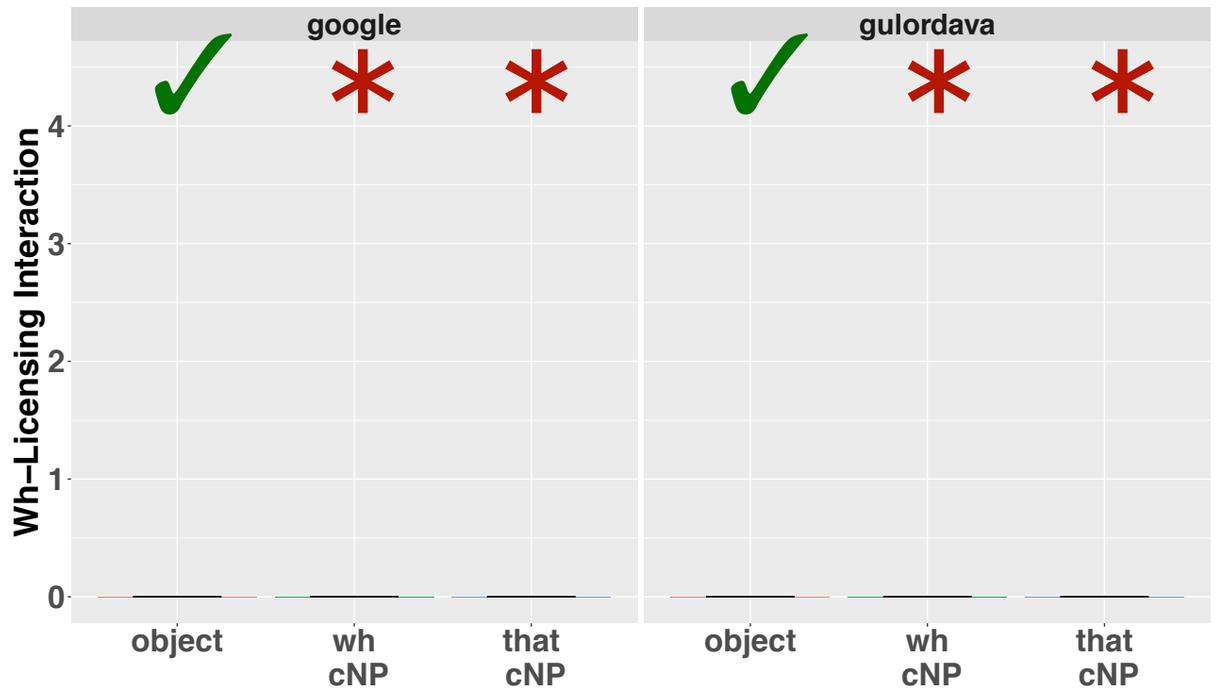



## Wh-Complimentizer Islands



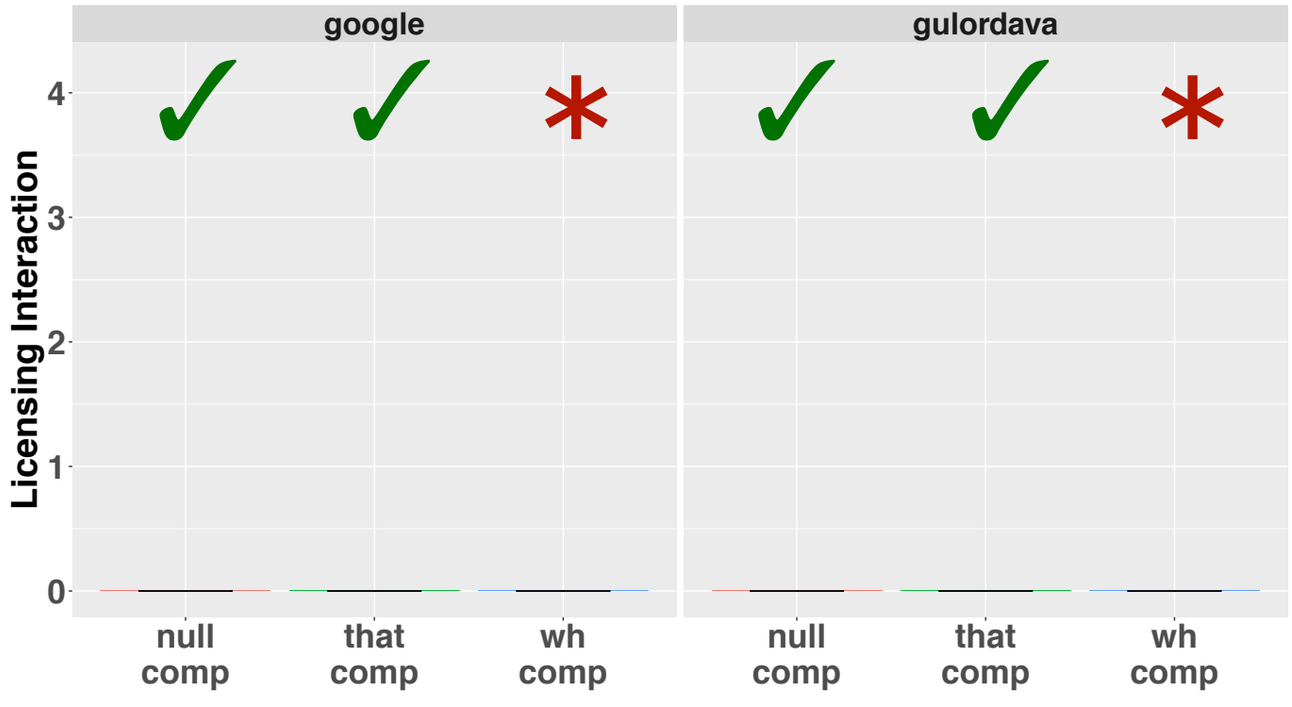
# Experiment 2-4: Results



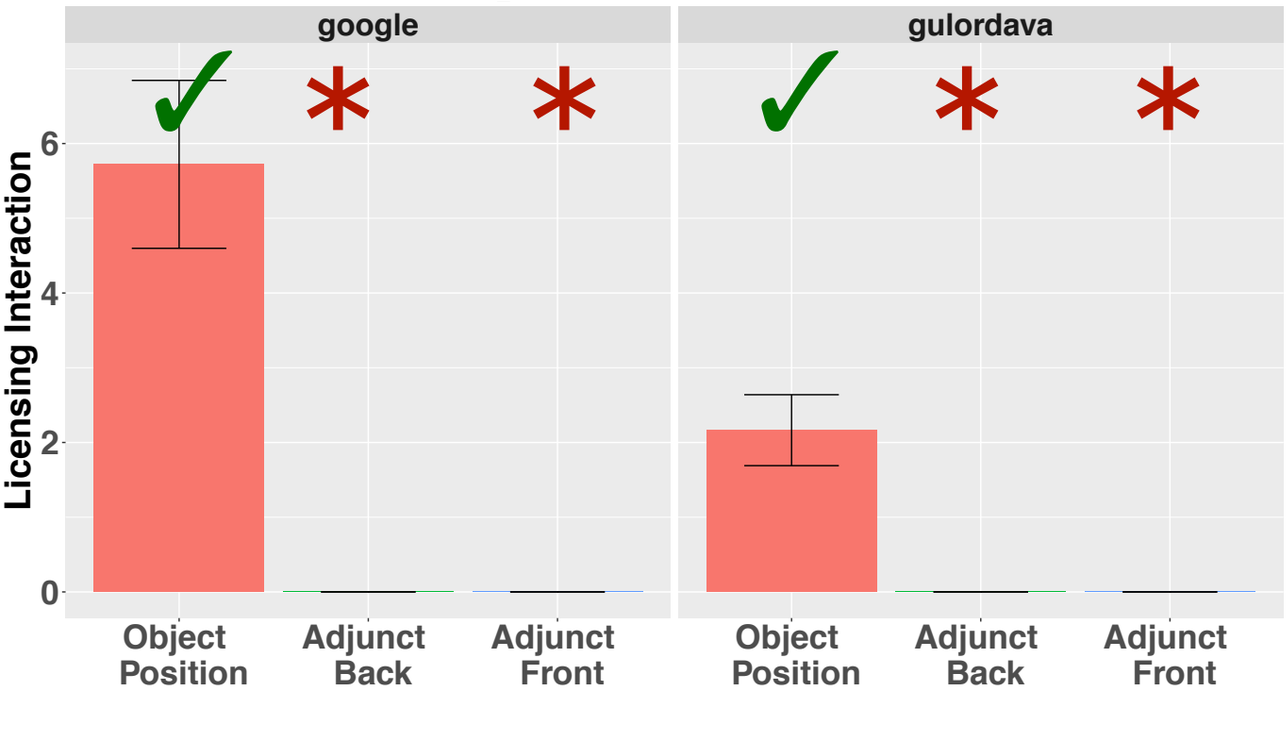

✓ grammatical, high licensing interaction

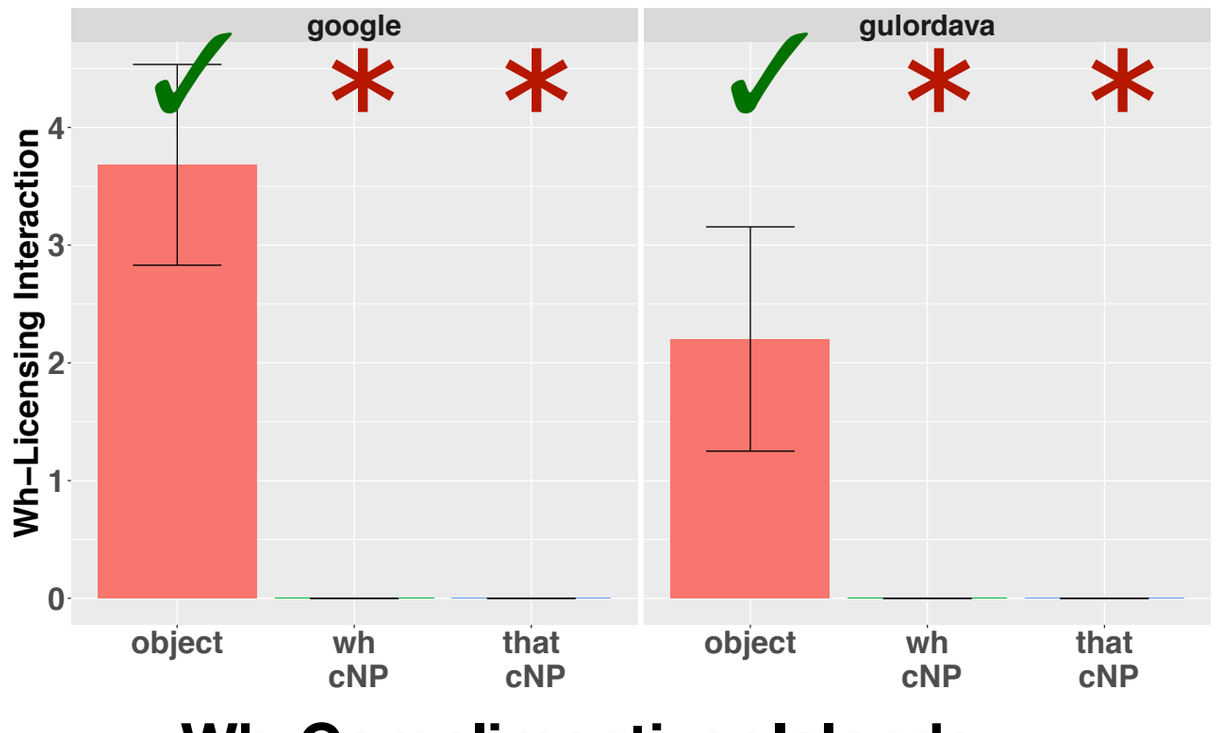
* island, low licensing interaction

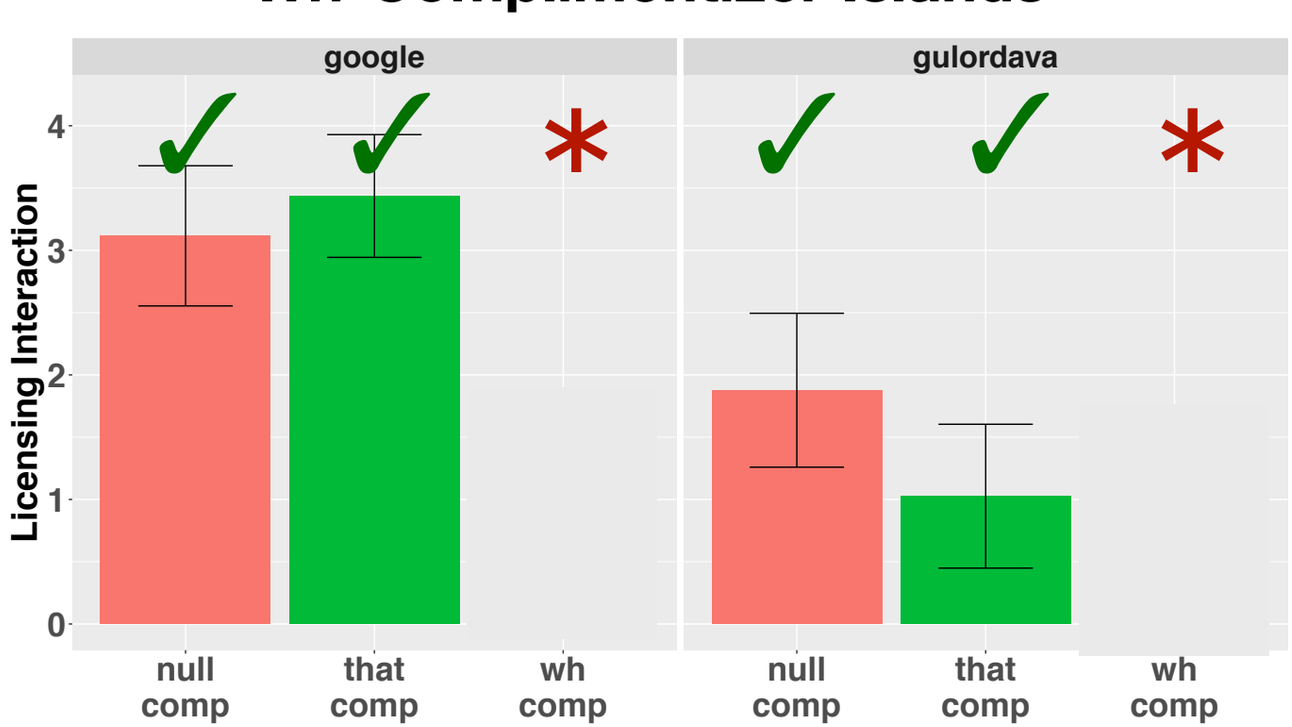


# Experiment 2-4: Results

## Grammatical Conditions: Strong Wh-Licensing Interaction

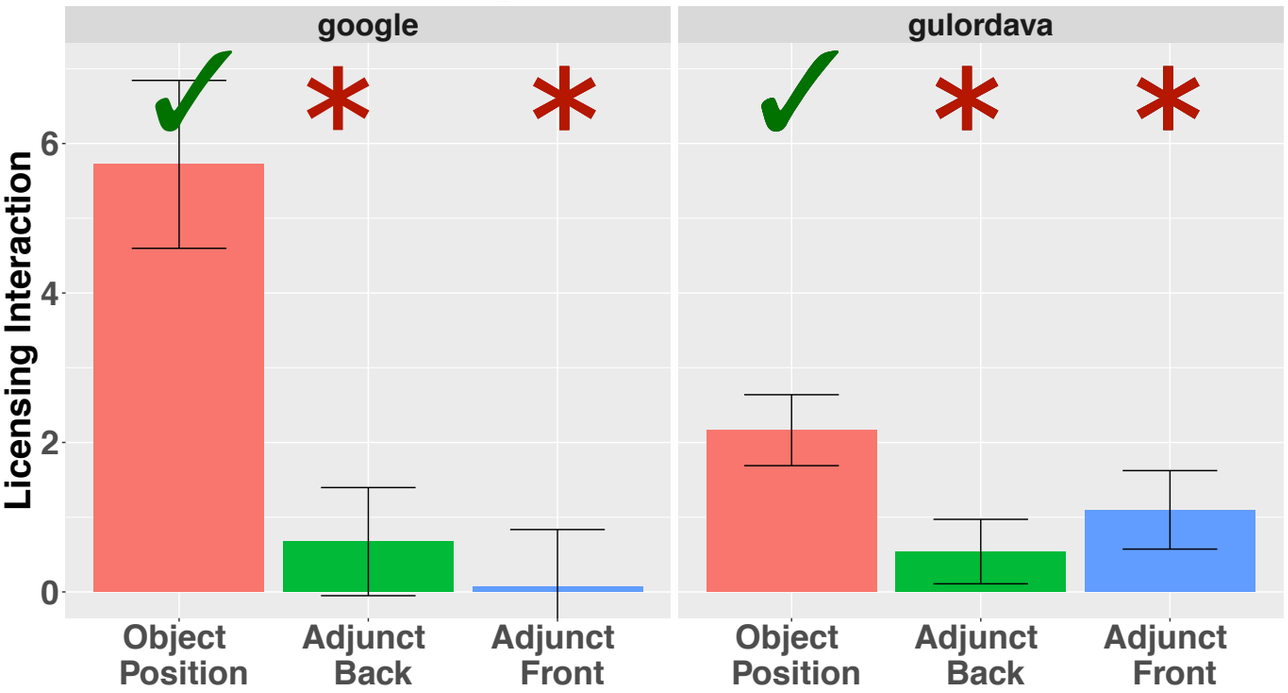


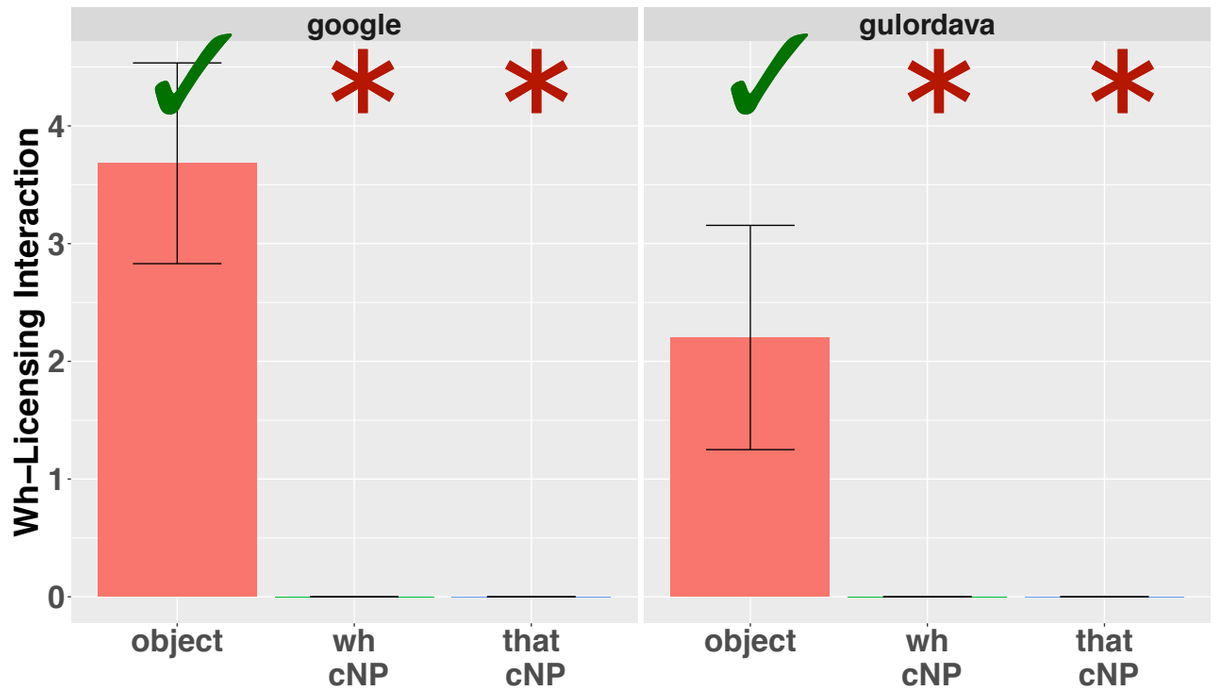



# Experiment 2-4: Results

## Adjunct Islands

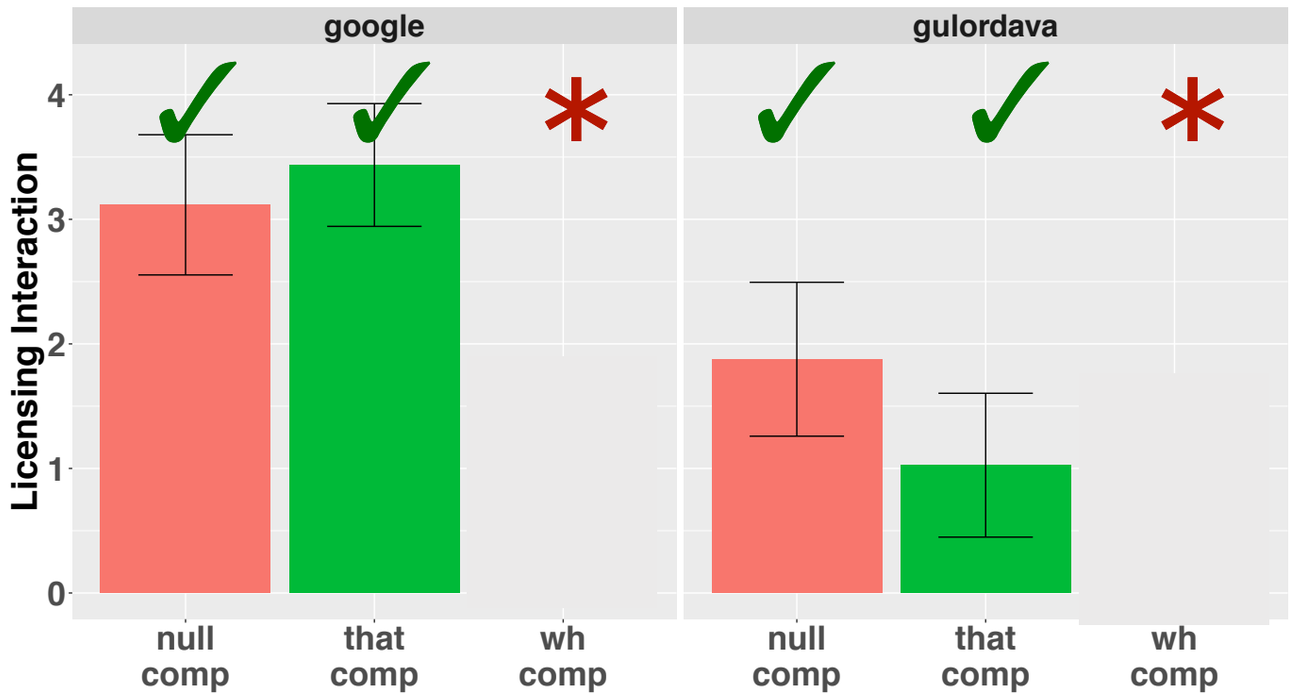



## Complex NP Islands



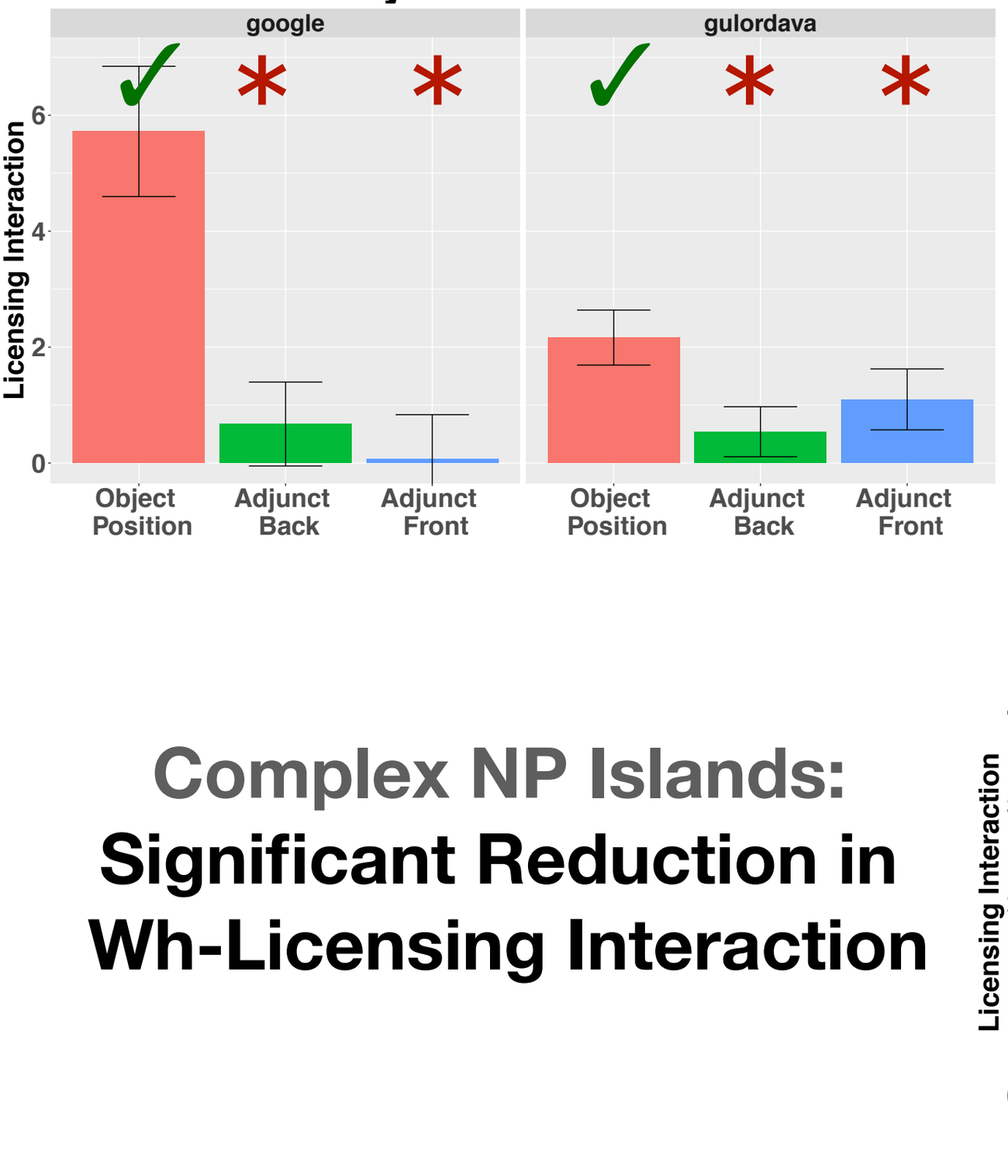

**Adjunct Islands:**
**Significant Reduction in Wh-Licensing Interaction**

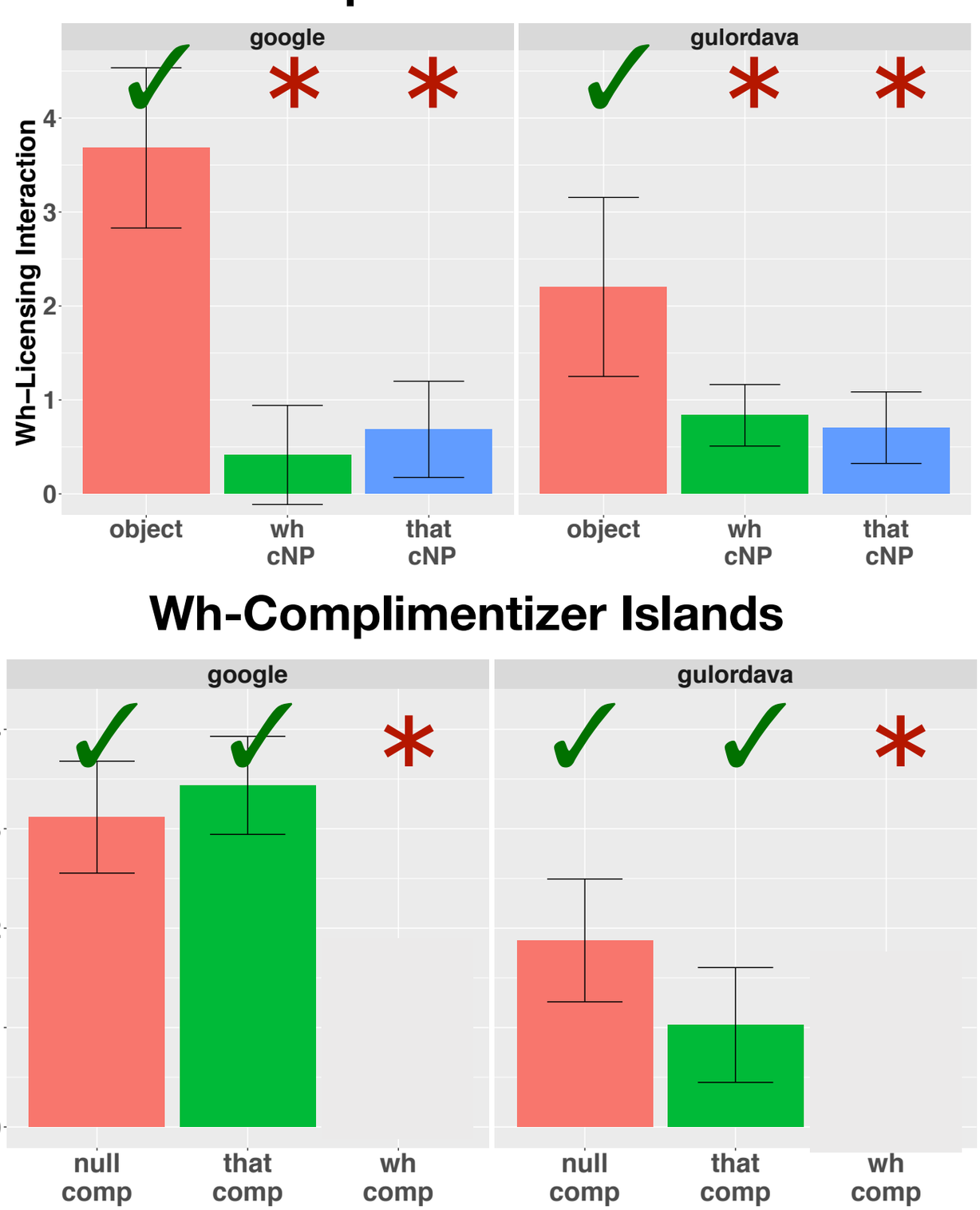
## Wh-Complimentizer Islands



# Experiment 2-4: Results

## Complex NP Islands:
## Significant Reduction in Wh-Licensing Interaction

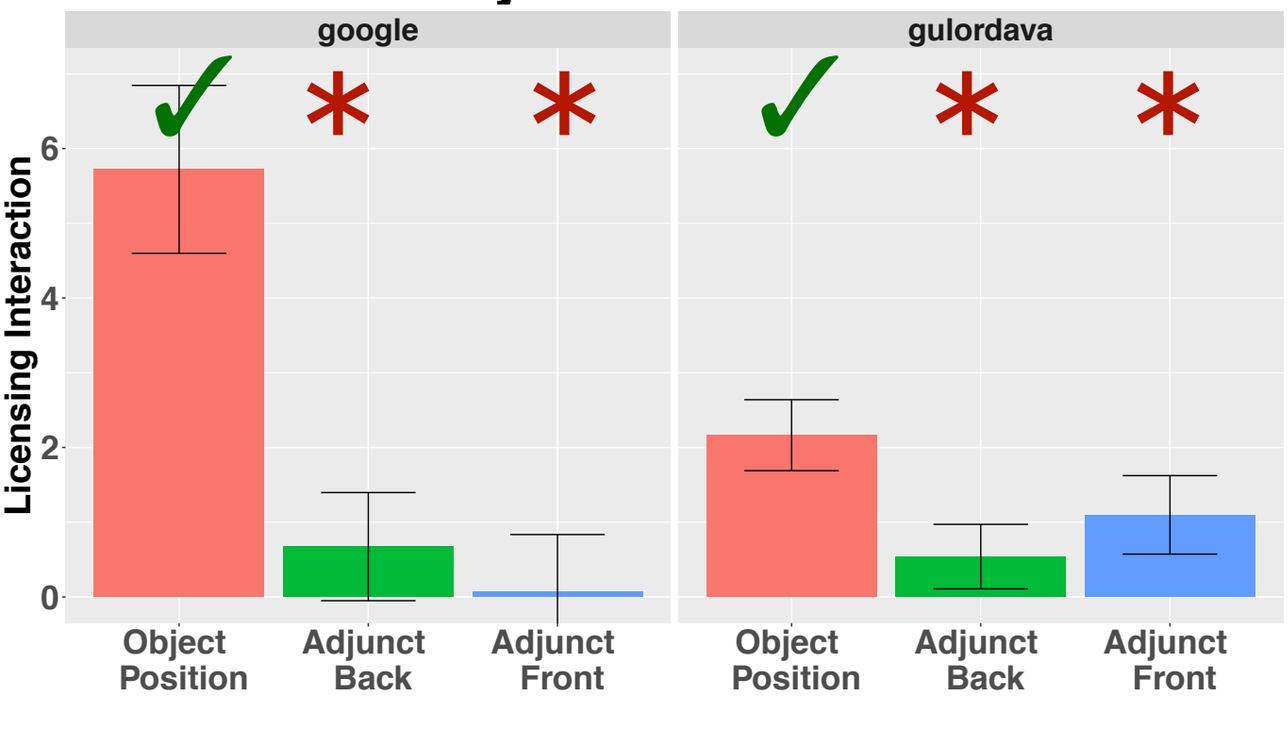


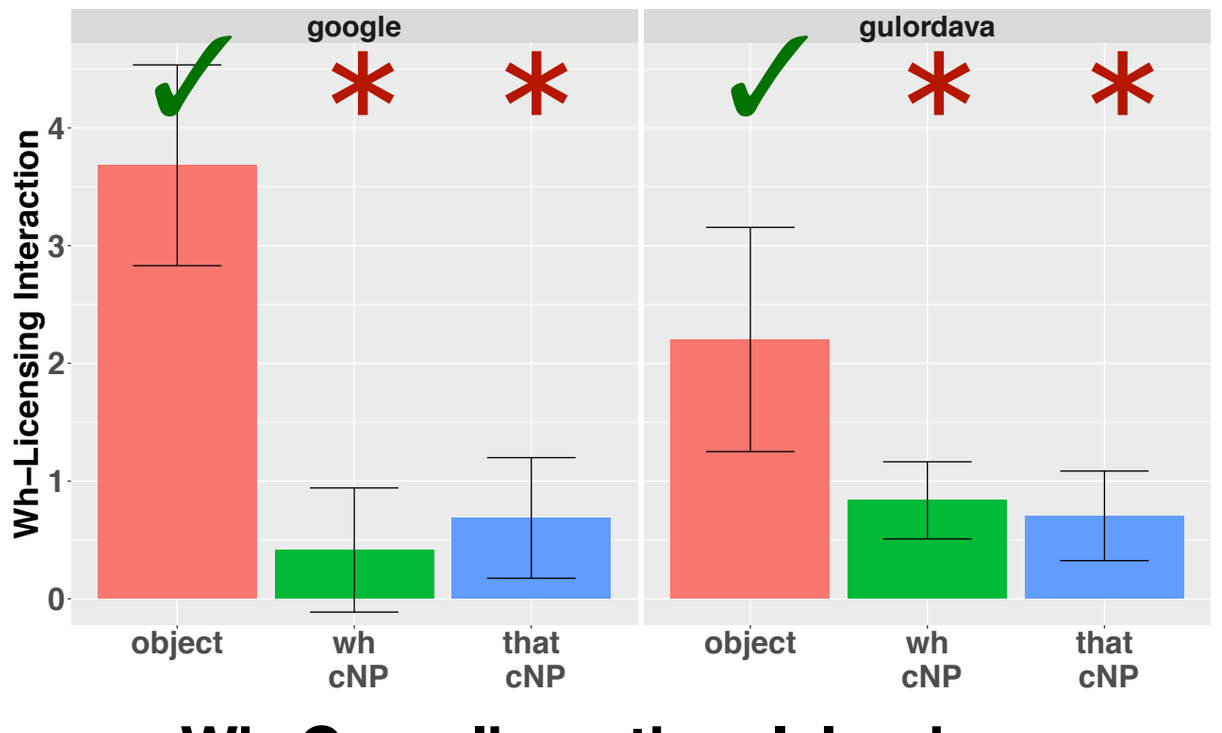



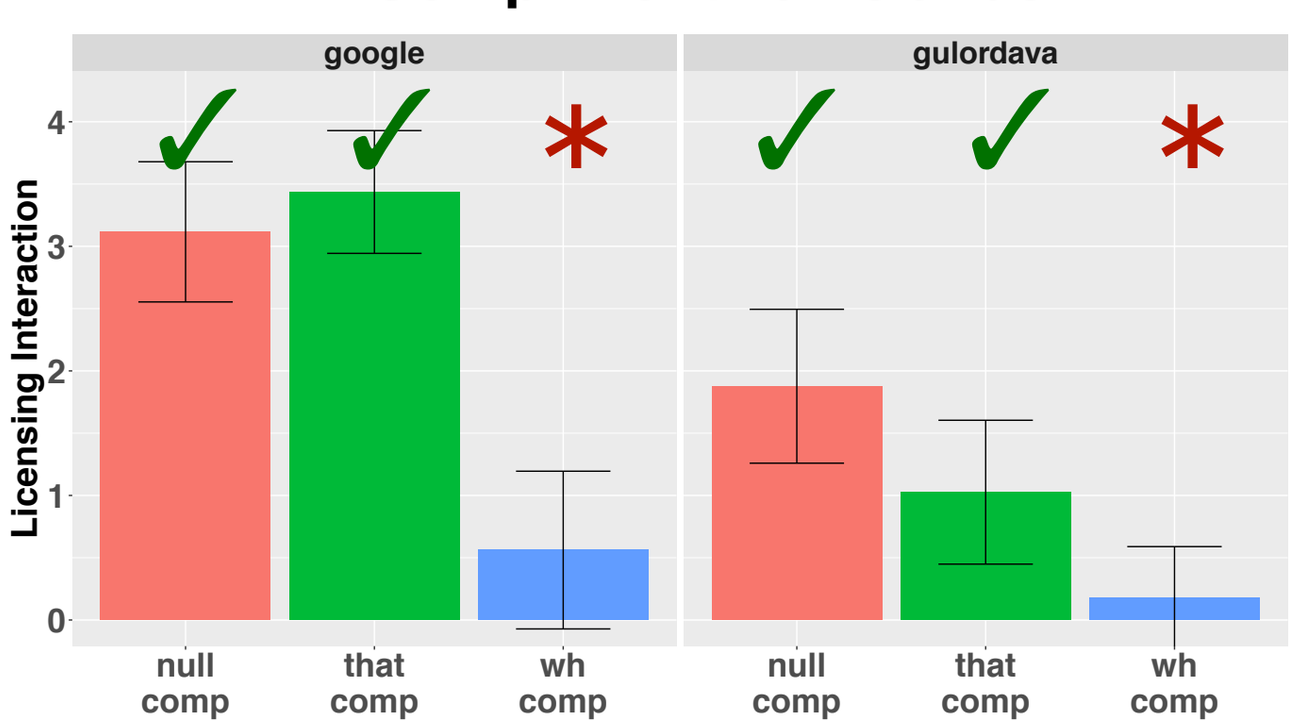
# Experiment 2-4: Results

## Wh-Islands:

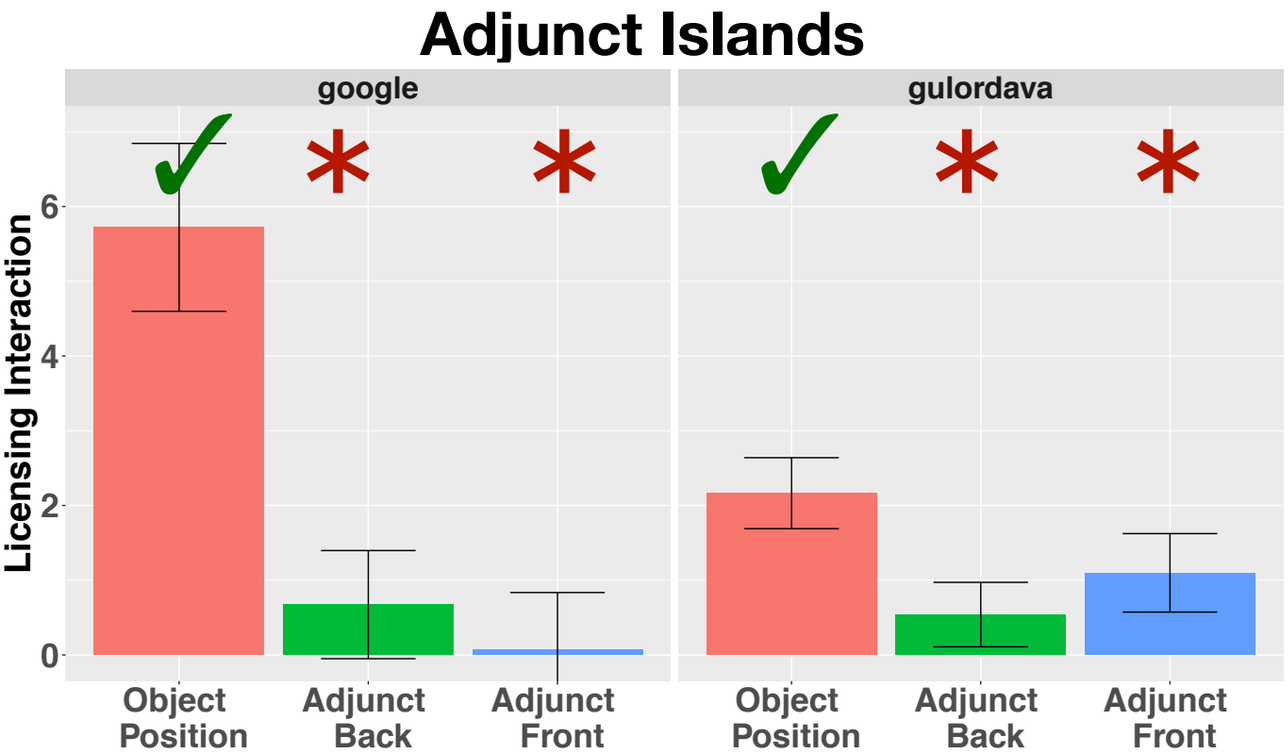
## Significant Reduction in Wh-Licensing Interaction

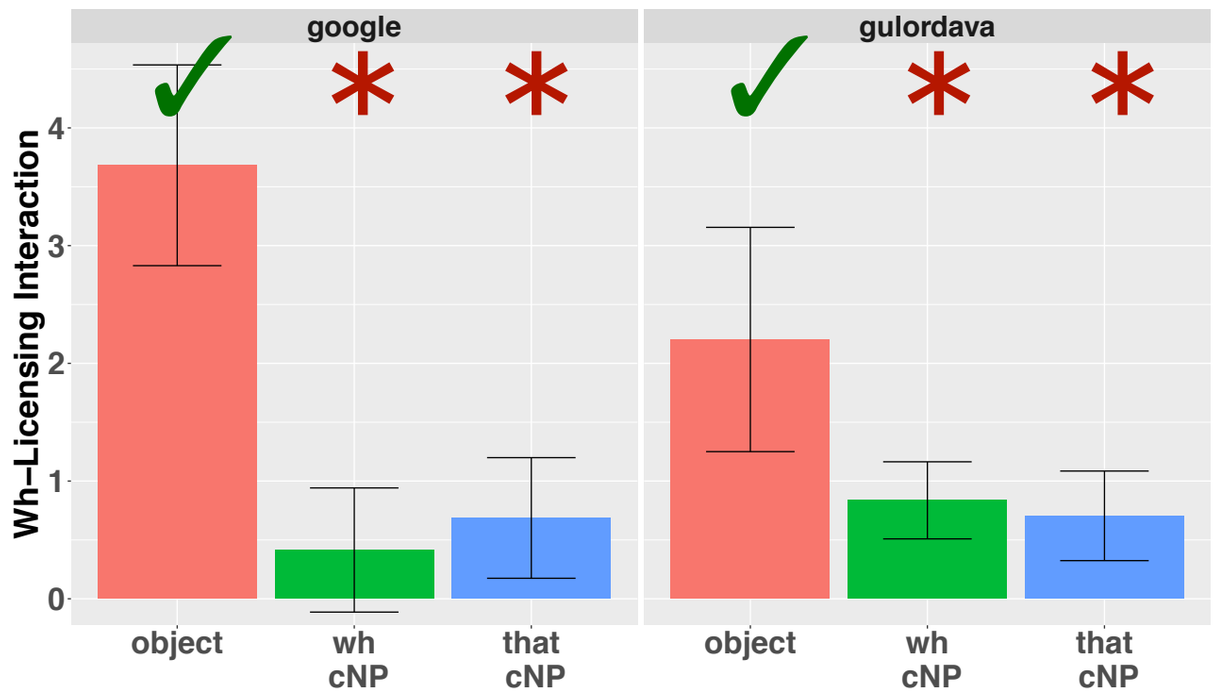


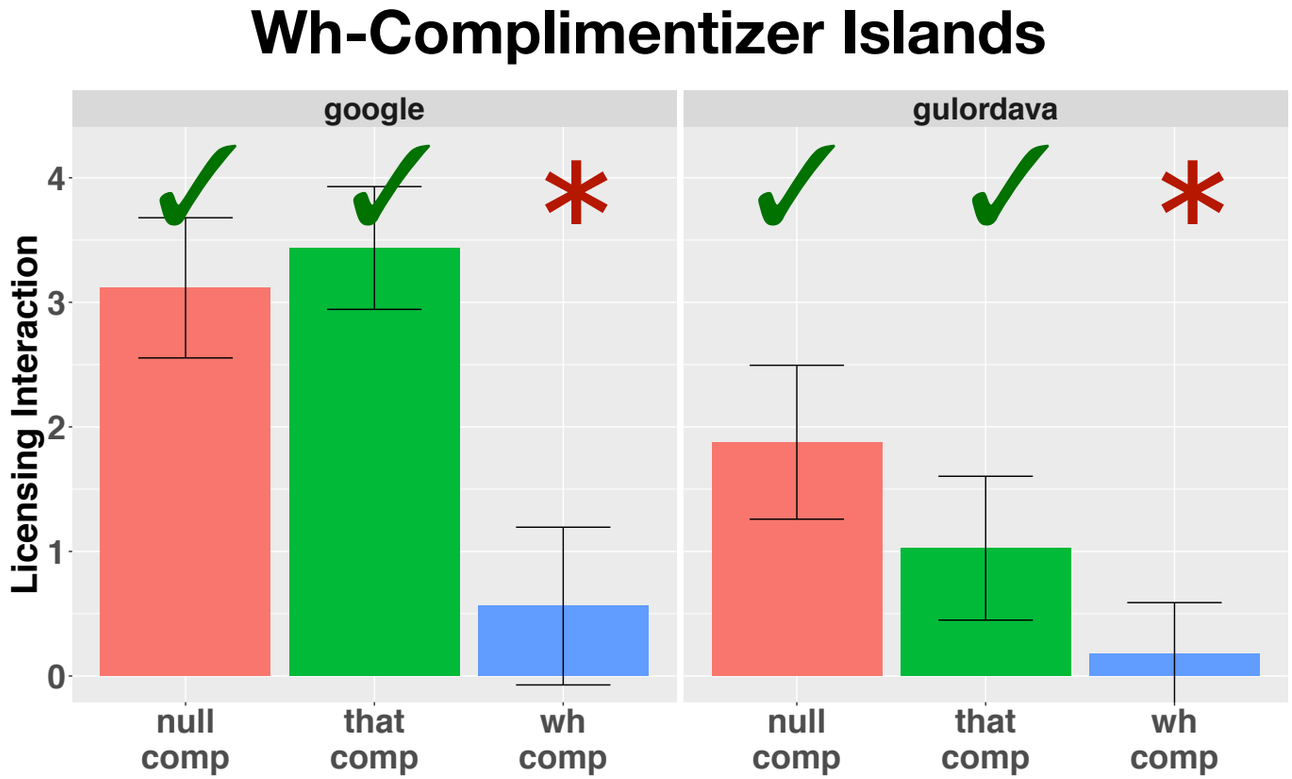



# Experiment 2-4: Results




**Both models,**
**All Conditions:**
**Attenuated Expectations**
**For gaps in Island Constructions**



# Experiment 5: Subject Islands

Filler—Gaps are licensed in PPs, but **not in subject position.**

*I know what…*

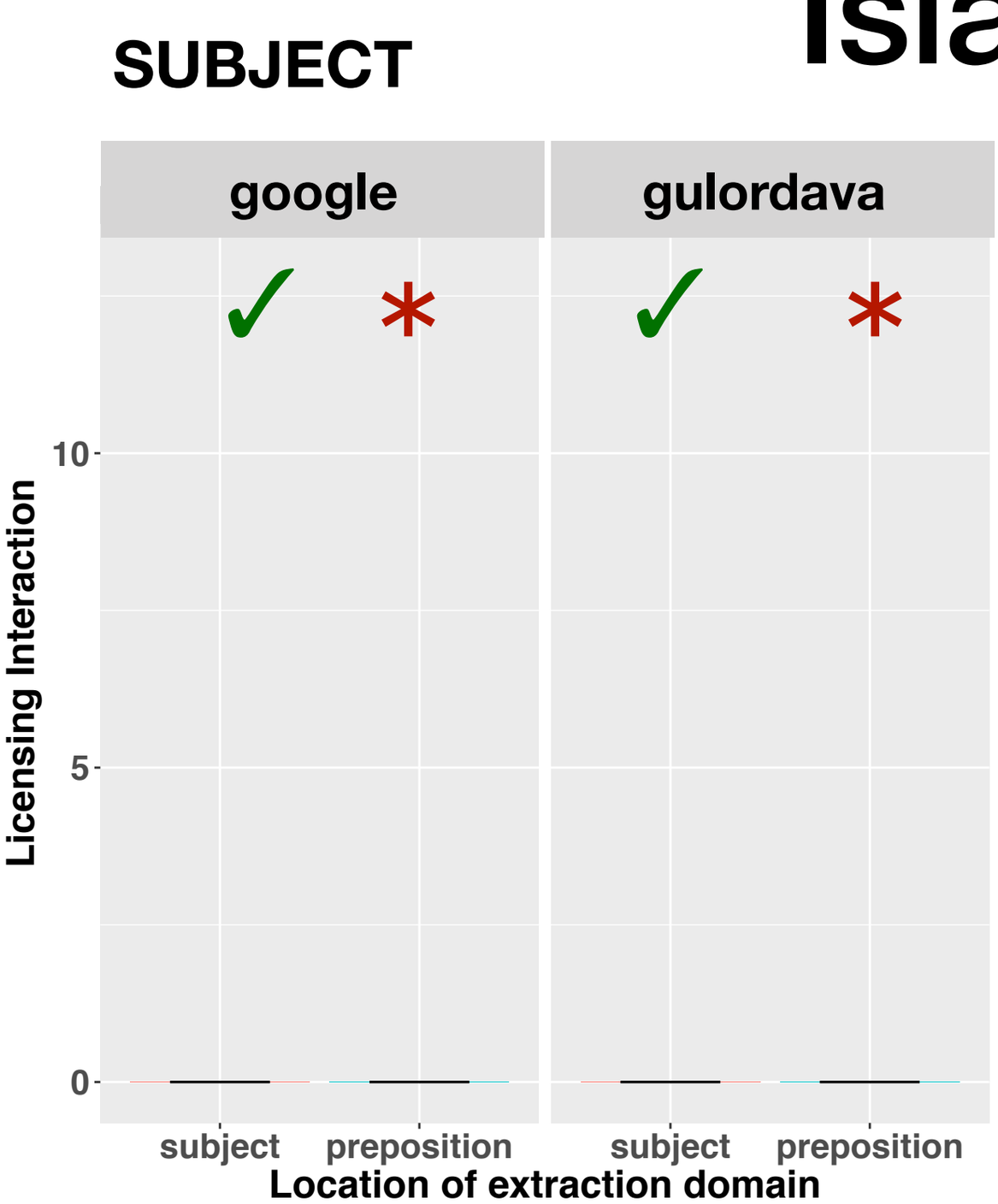
✓ *…the family saw ___ in the museum.*
*[Object]*

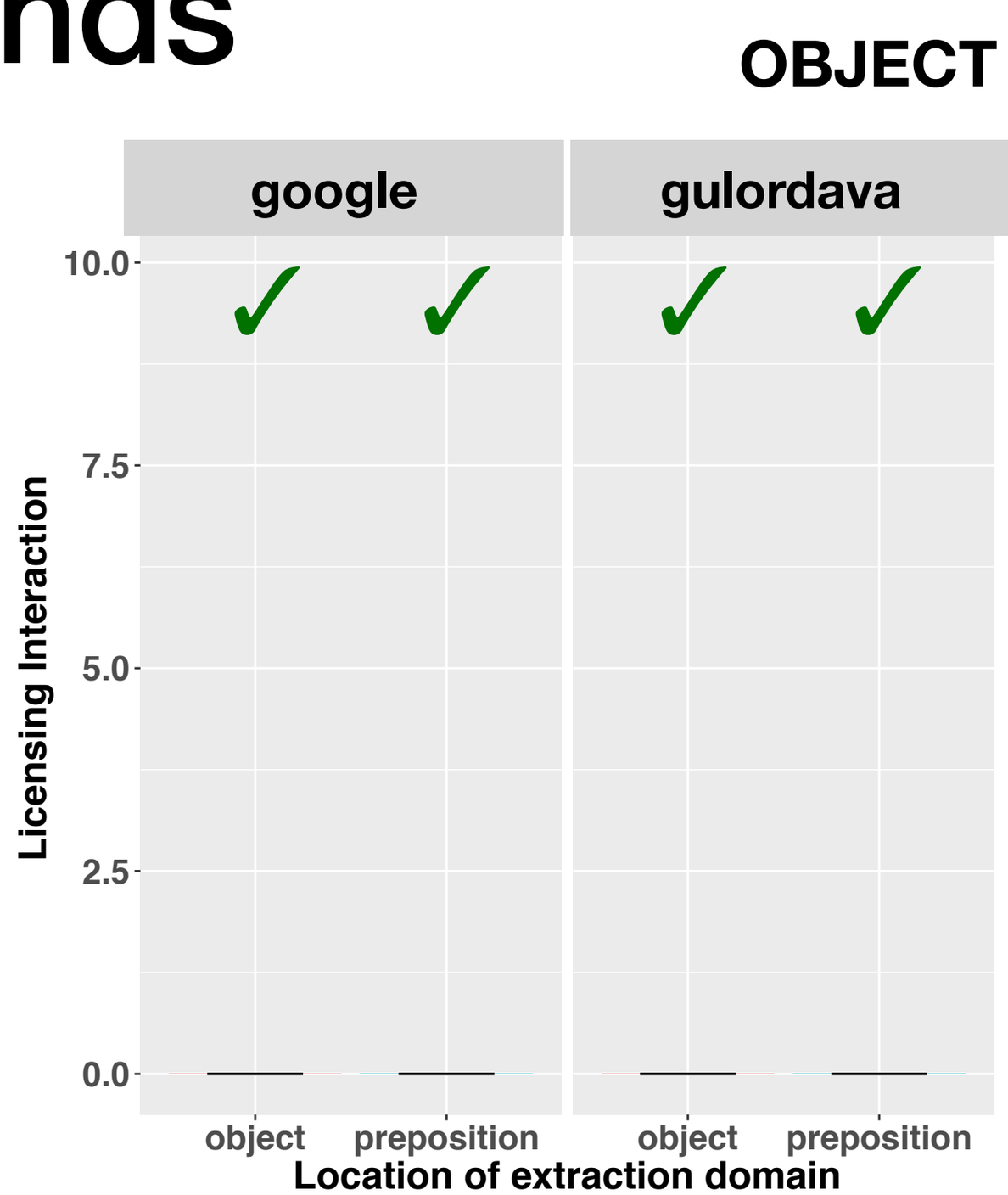
✓ *…the family saw a painting of ___ in the museum. [PP Object]*

✓ *… ___ fetched a high price at auction.*
*[Subject]*

* *…a painting of ___ fetched a high price at auction. [PP Subject]*

# Experiment 5: Subject Islands

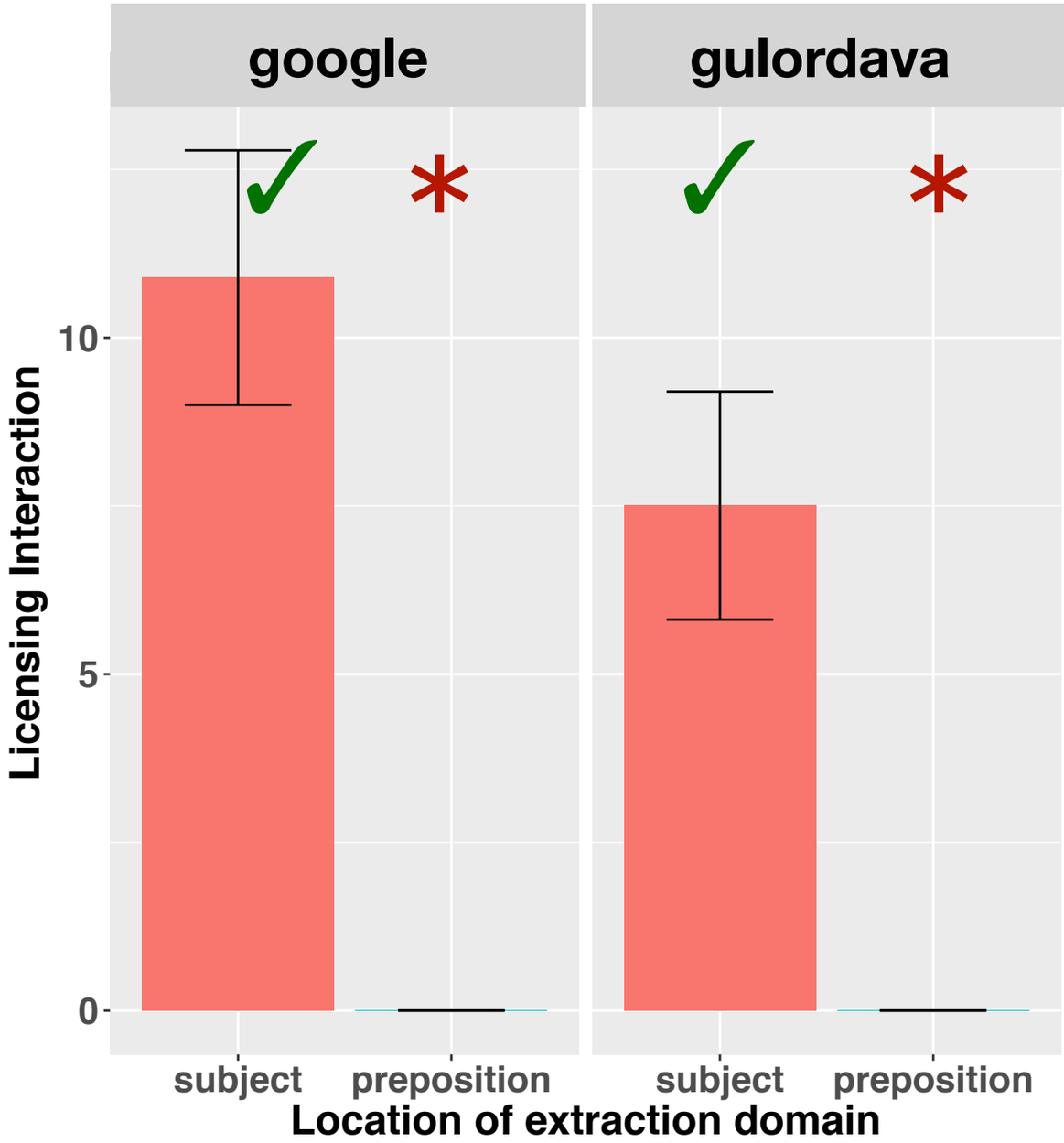


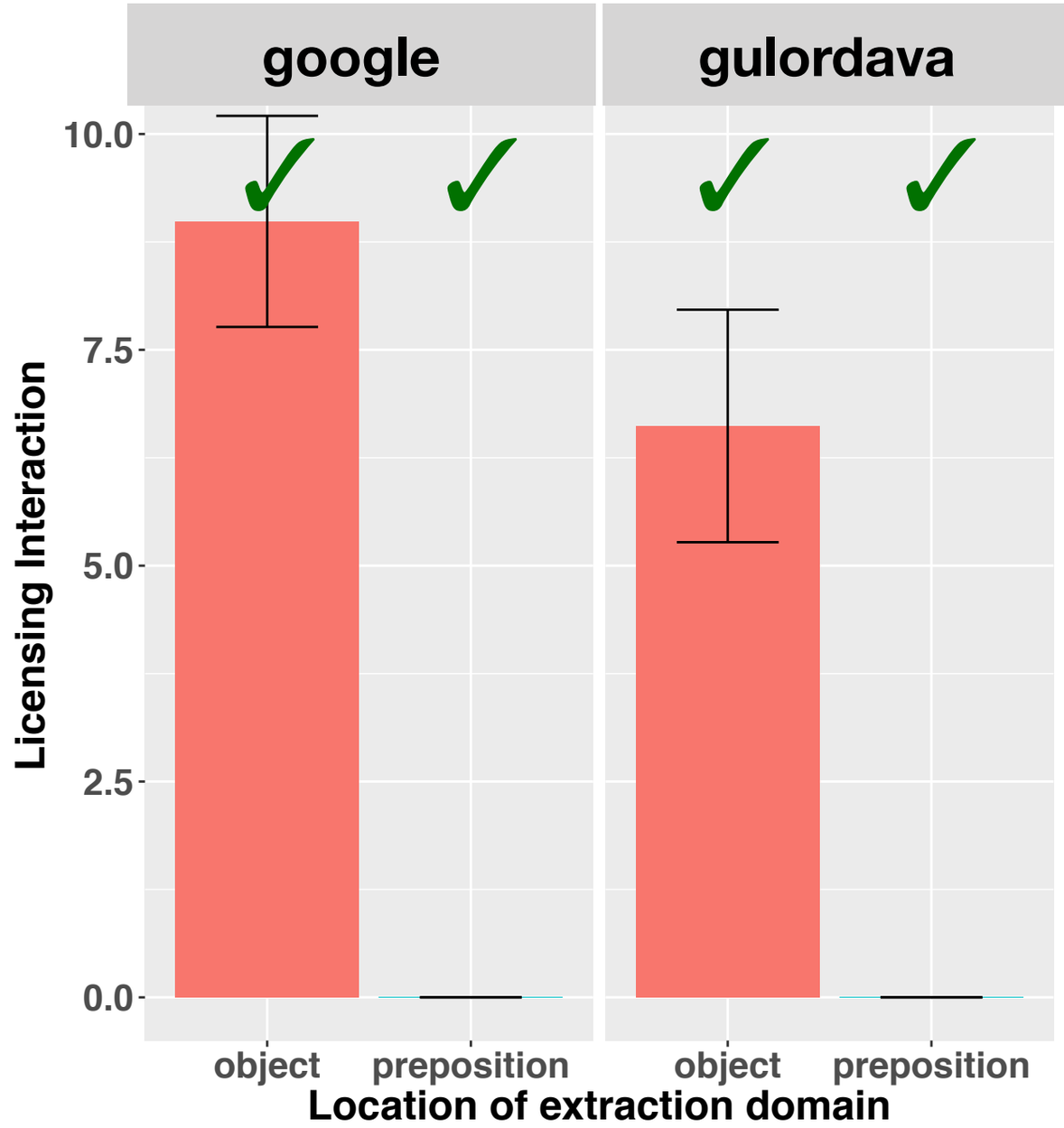



# Experiment 5: Subject Islands

SUBJECT

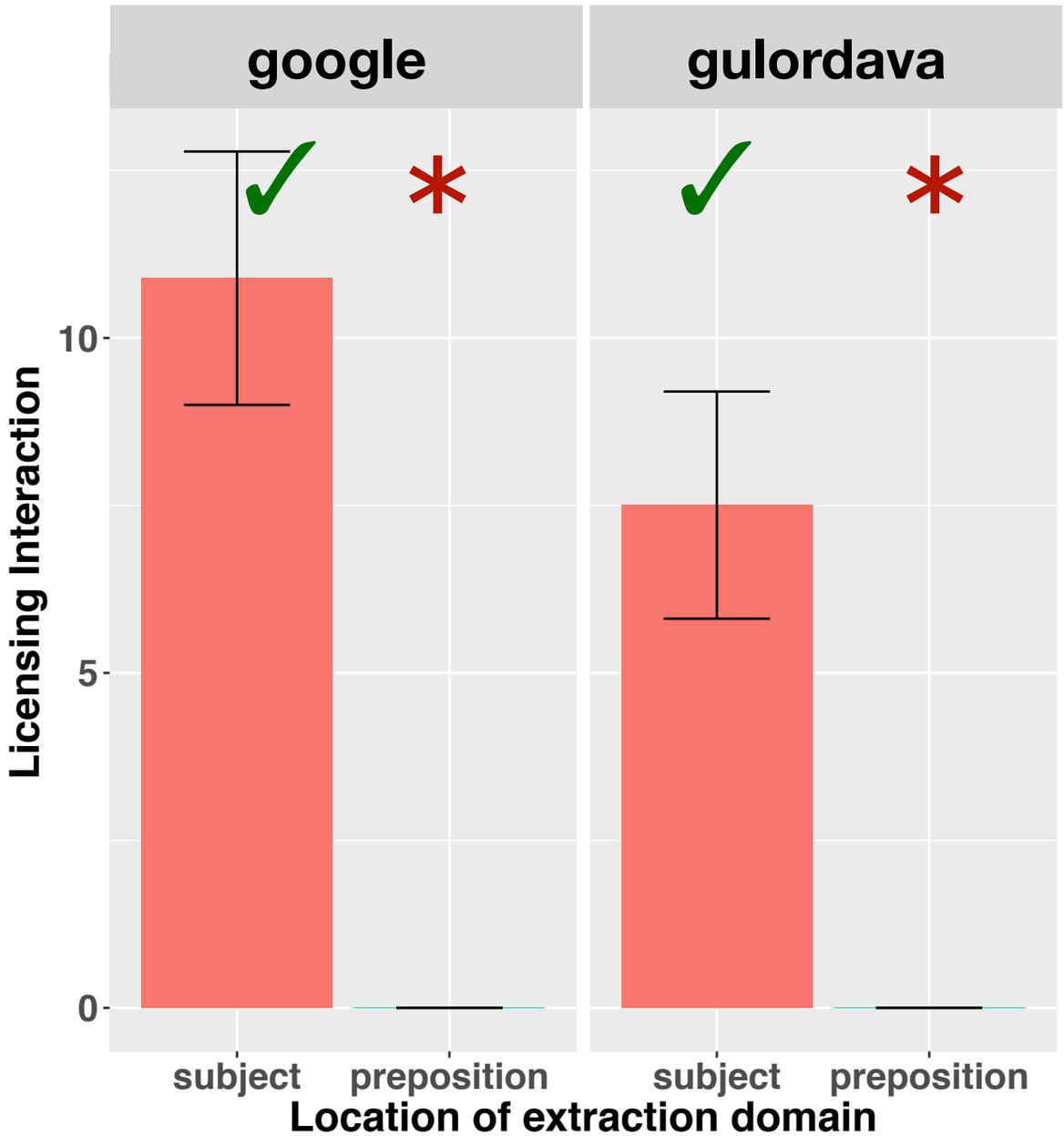
OBJECT



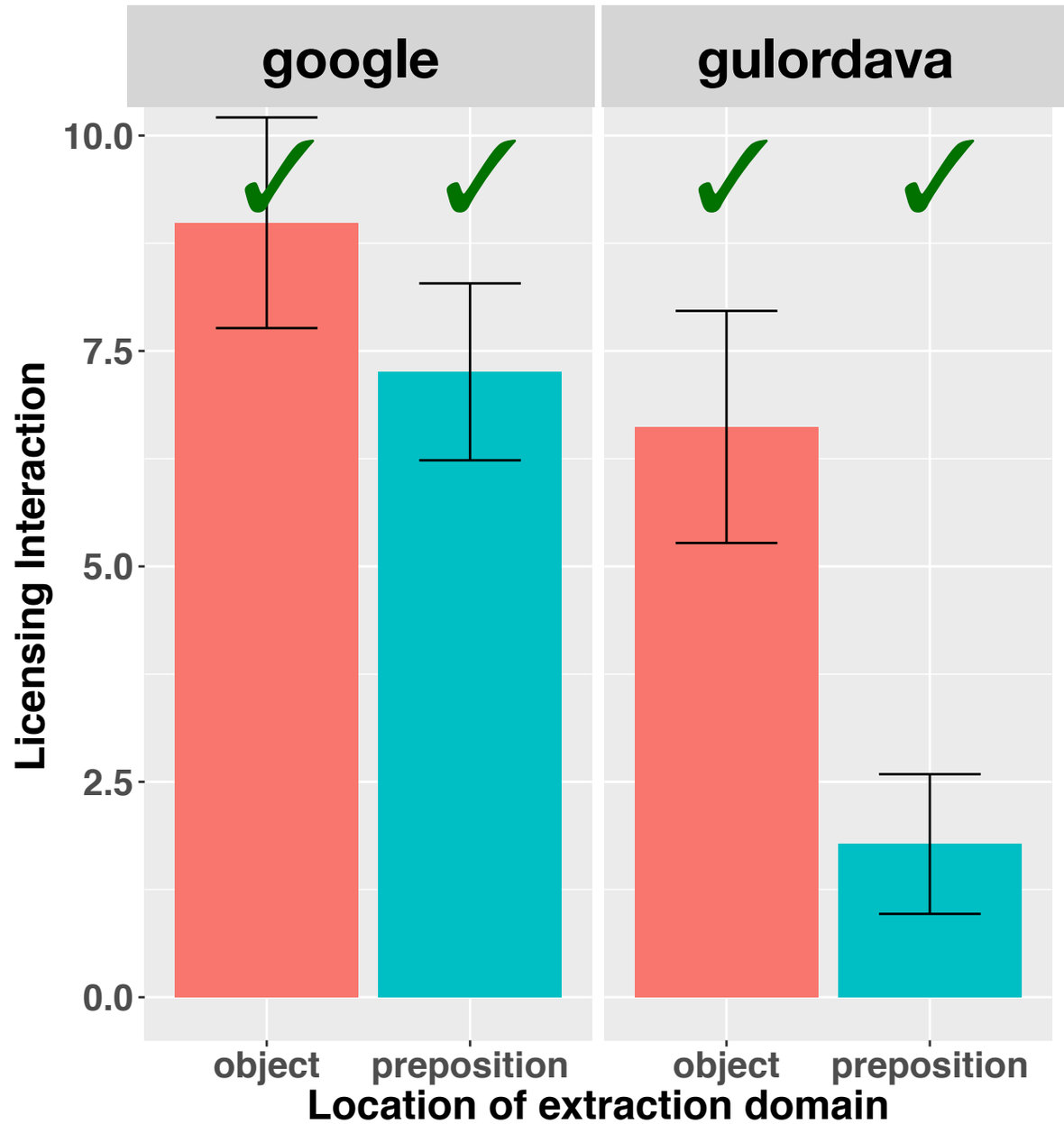



# Experiment 5: Subject Islands

SUBJECT

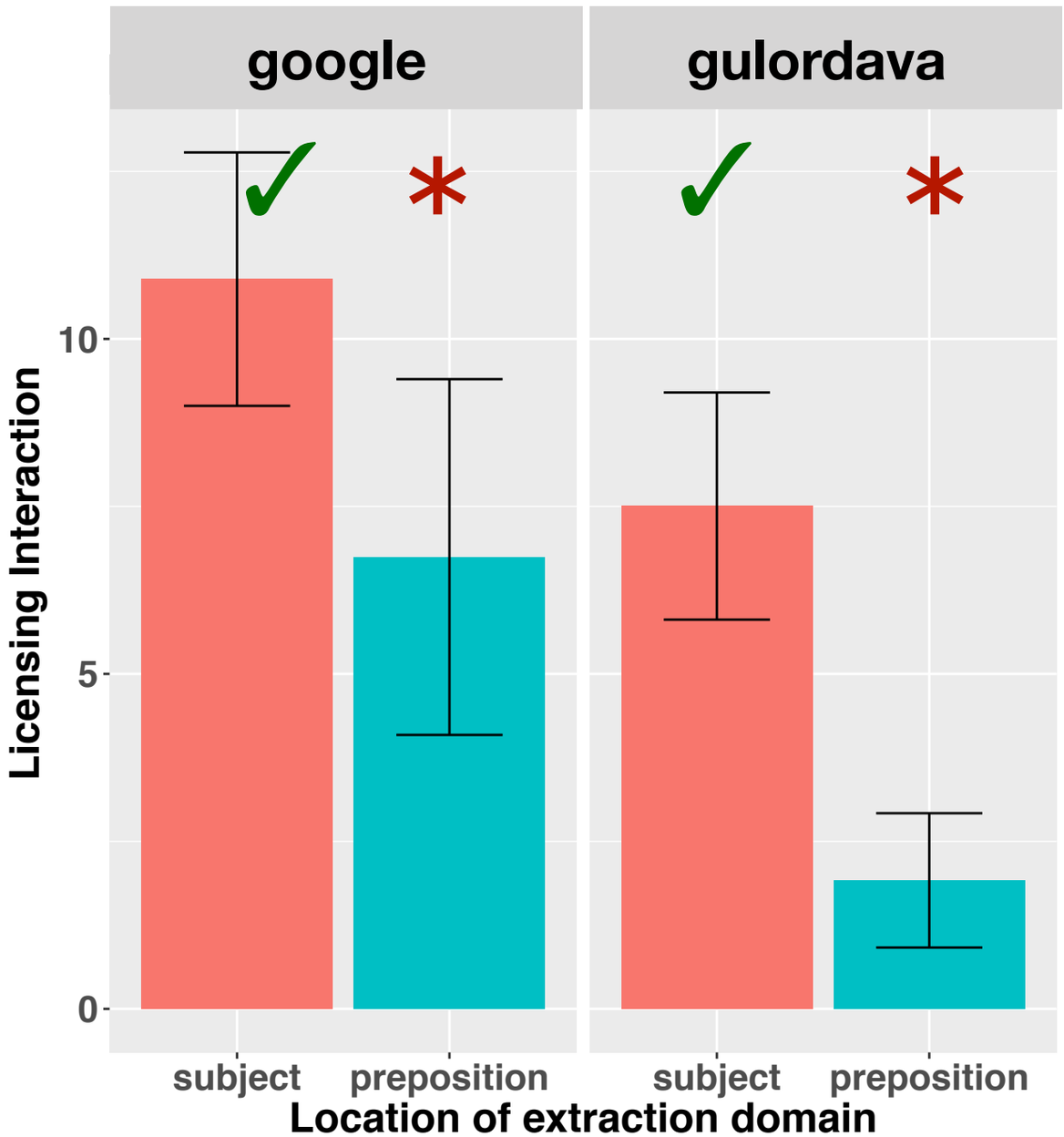



OBJECT

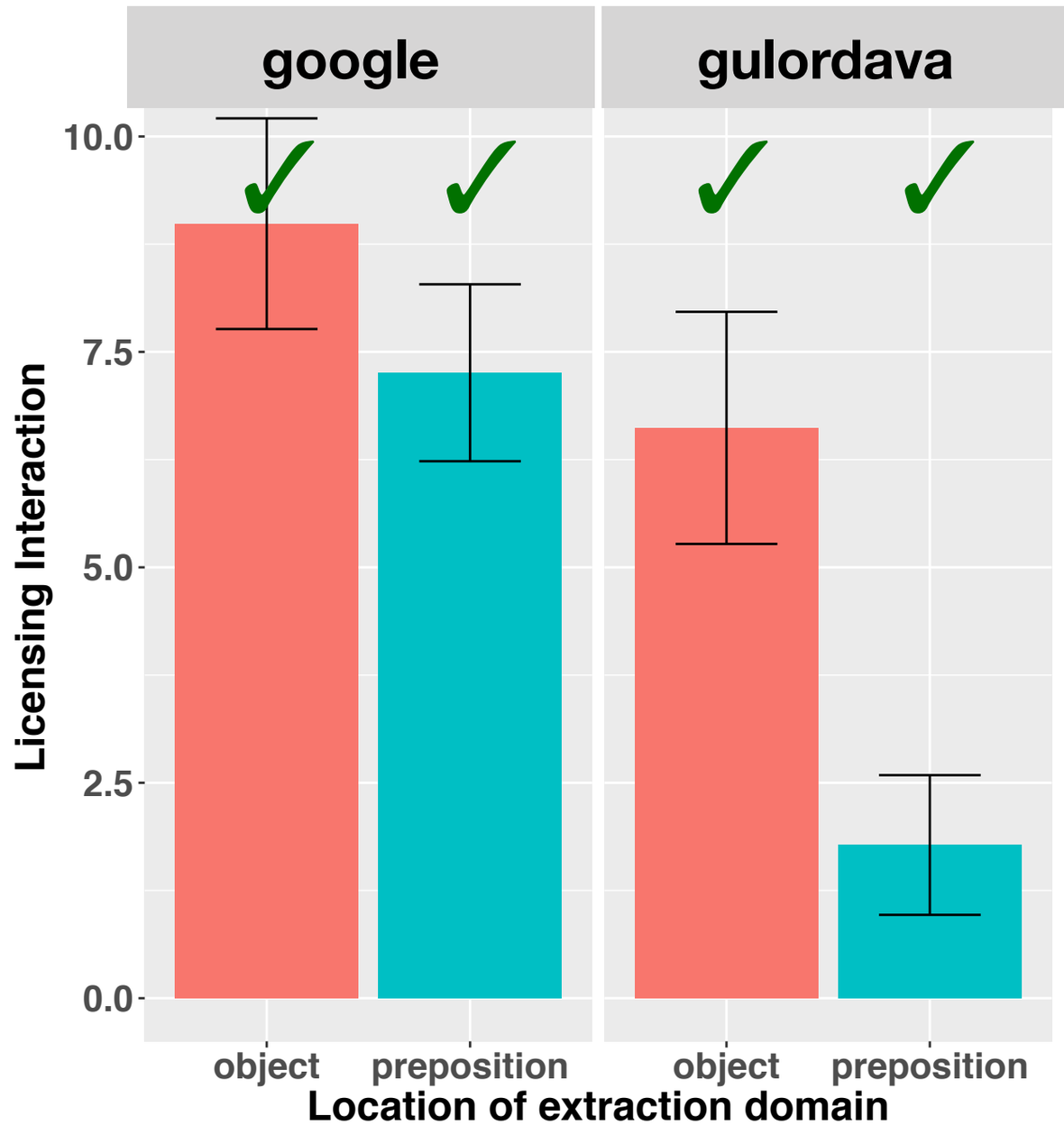


# Experiment 5: Subject Islands

SUBJECT

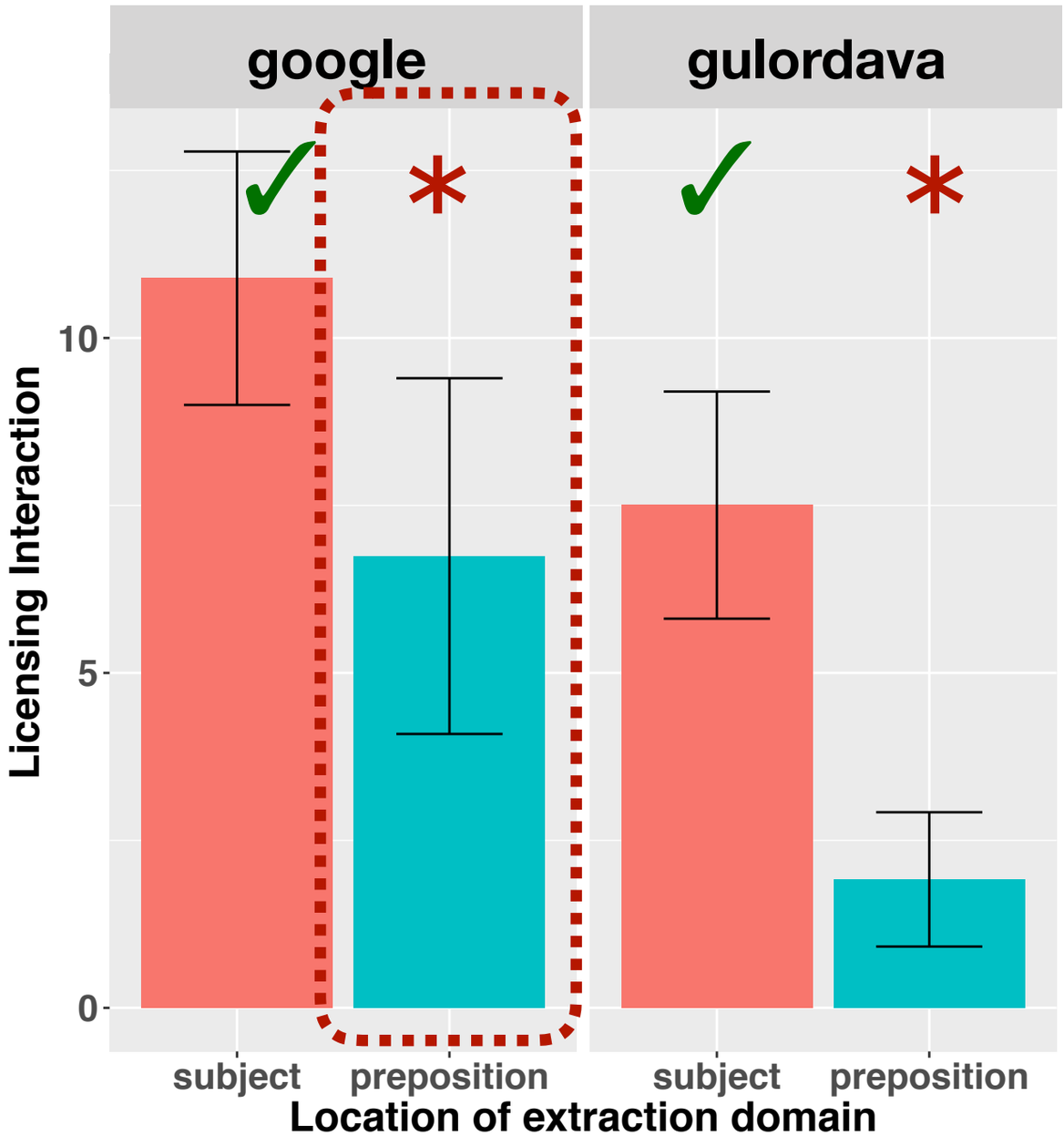



OBJECT

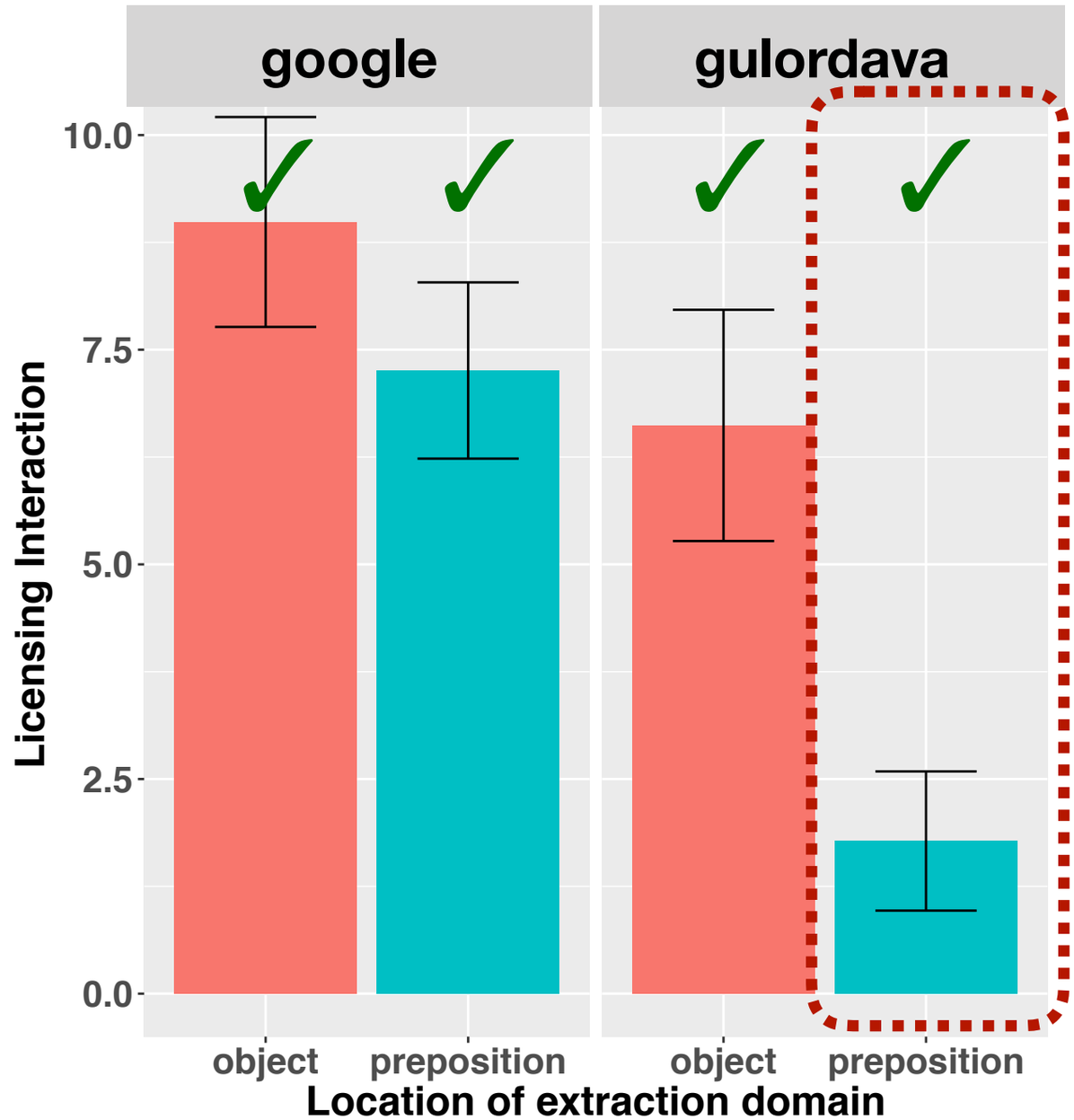


# Experiment 5: Subject Islands

SUBJECT



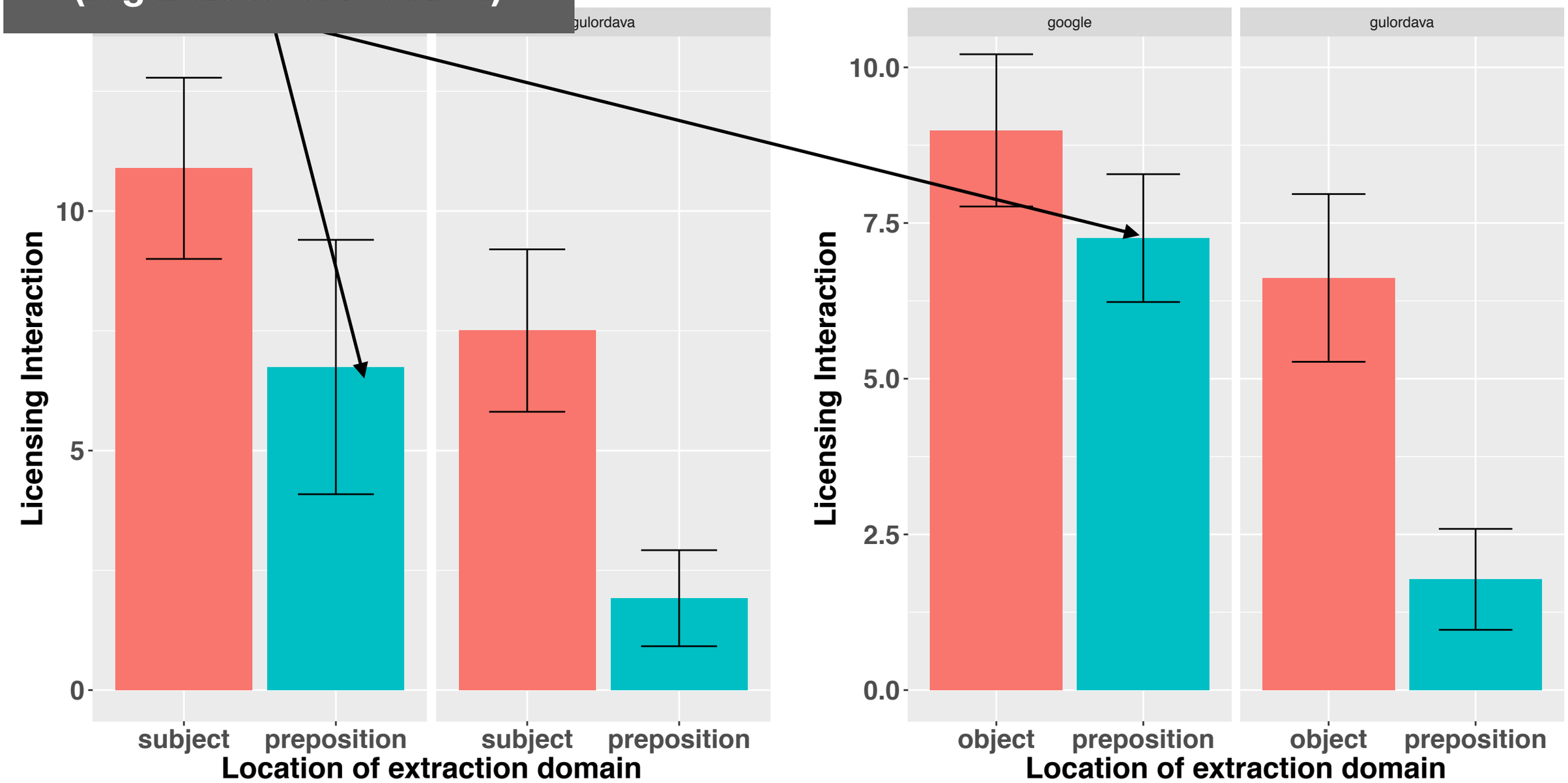

OBJECT



# Experiment 5: Subject Islands

**Google Model: Gaps are always okay in PPs (regardless of location)**

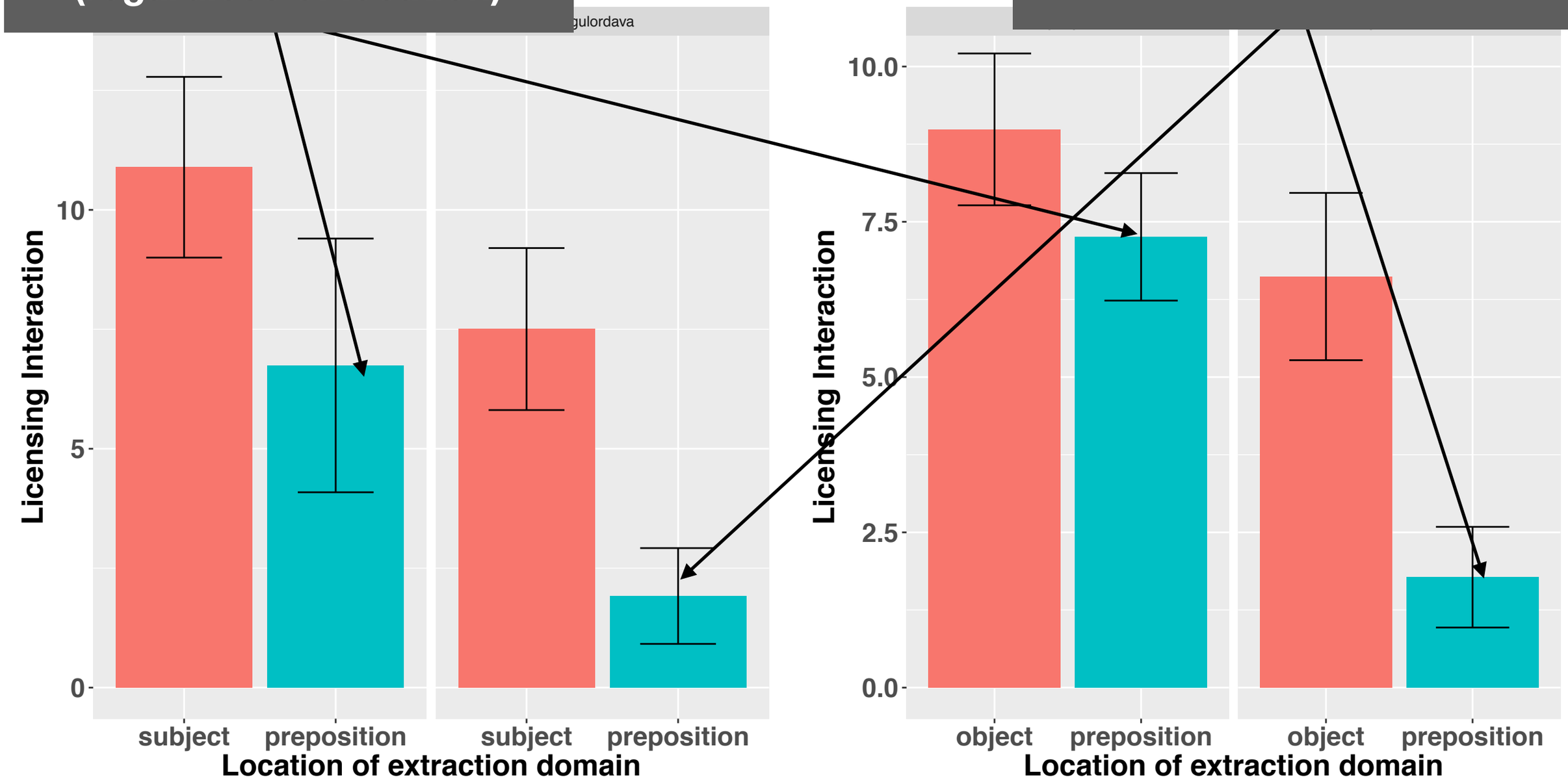
**OBJECT**



# Experiment 5: Subject Islands

**Google Model: Gaps are always okay in PPs (regardless of location)**

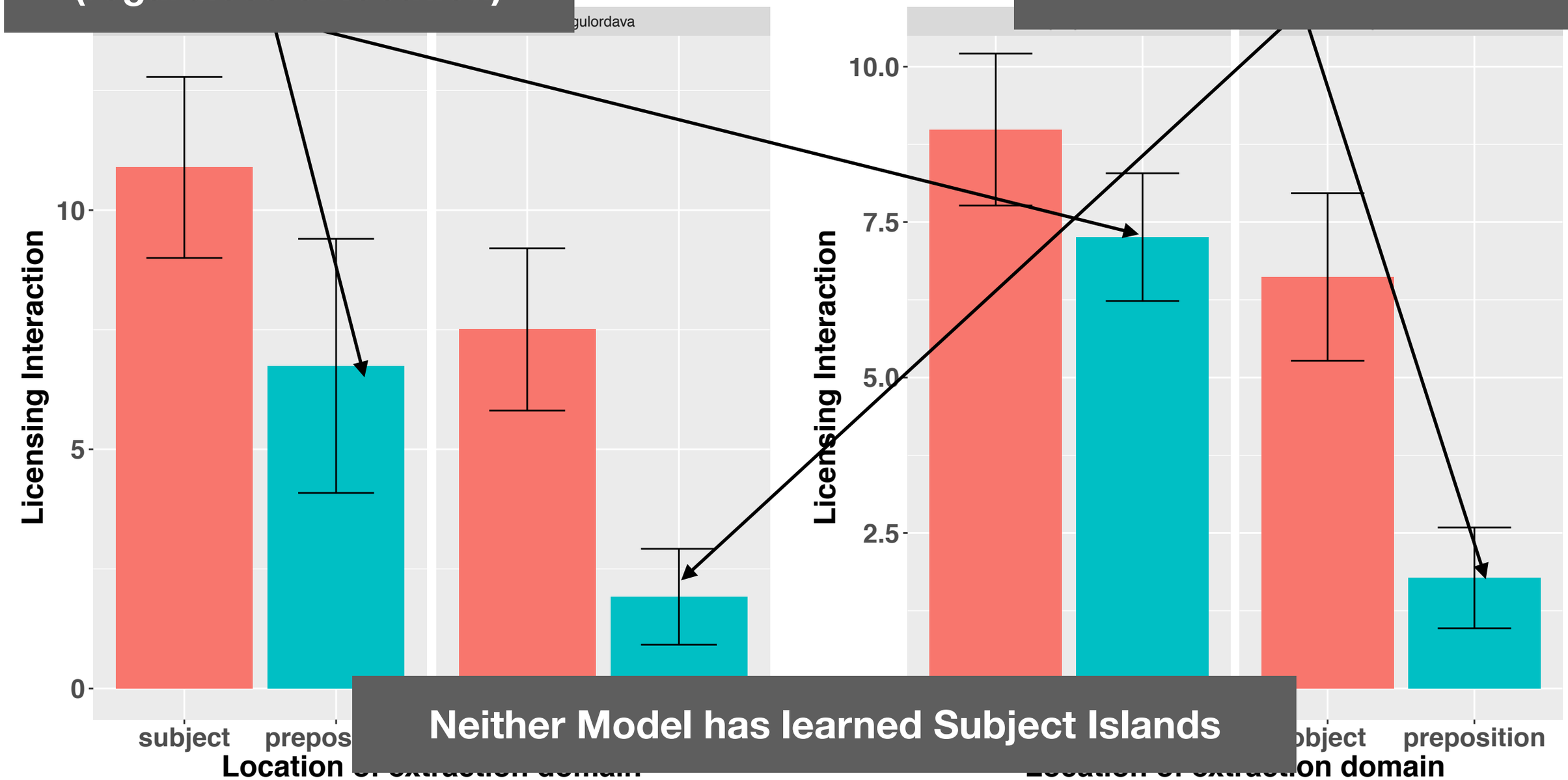
**Gulordava Model: Gaps are never okay in PPs Regardless of location)**



# Experiment 5: Subject Islands

**Google Model: Gaps are always okay in PPs (regardless of location)**

**Gulordava Model: Gaps are never okay in PPs Regardless of location)**




**Neither Model has learned Subject Islands**

# Islands: Review

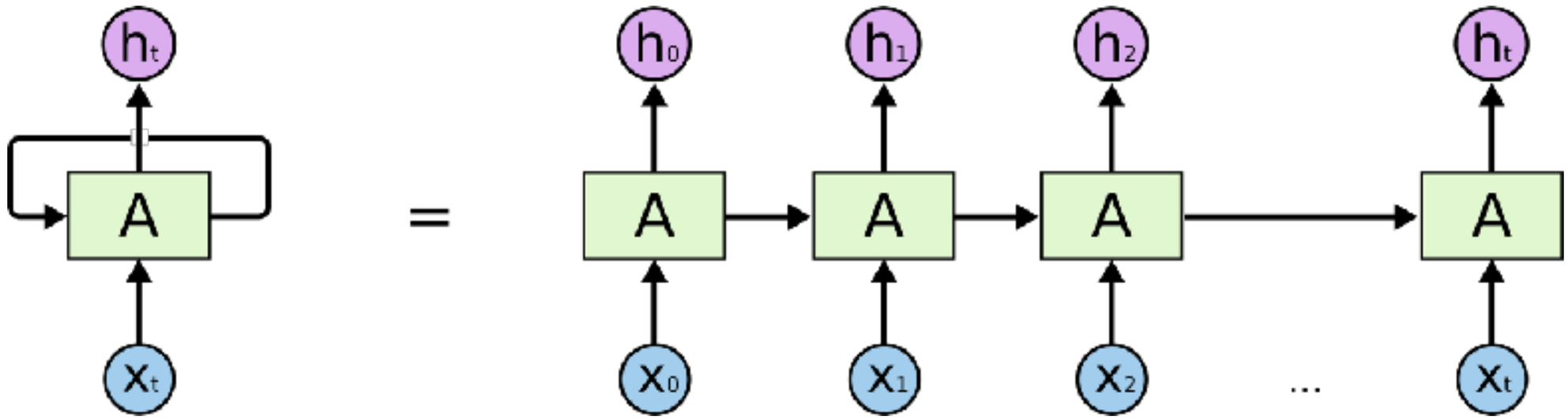
| | Google Model | Gulordava Model |
|---|---|---|
| **Adjunct Islands** | Yes | Yes |
| **Wh-Islands** | Yes | Gradiently |
| **Complex NP Islands** | Yes | Yes |
| **Subject Islands** | No | No |

# Islands: Preview

| | Google Model | Gulordava Model |
|---|---|---|
| Adjunct Islands | Yes | Yes |
| Wh-Islands | Yes | Gradiently |
| Complex NP Islands | Yes | Yes |
| Subject Islands | No | No |
| Sentential Subject Islands | No | No |
| Left-Branch Islands | Yes | Yes |
| Coordination Islands | Gradiently | Gradiently |

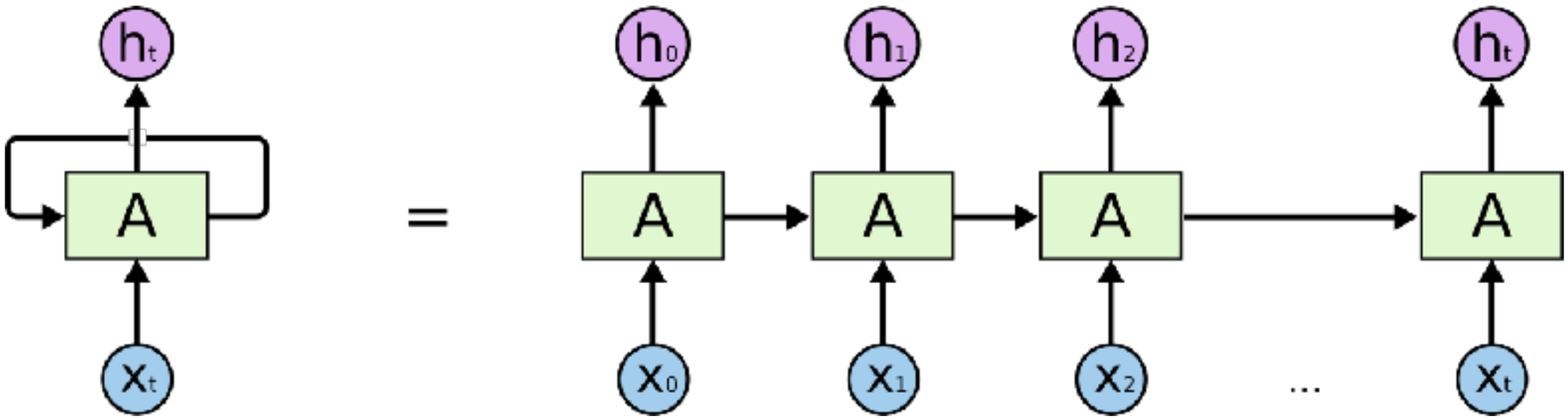
# In Conclusion

- RNN LMs learn the **filler-gap dependency**.
- RNN LMs  learn some of the **island constraints**.



# In Conclusion

- RNN LMs learn the **filler-gap dependency**.
- RNN LMs  learn some of the **island constraints**.




**Thank you for listening!**

# Extra Slides!

# Experiment 12: Coordination Islands

Gaps cannot occur in **one half** of a coordinated construction

*I know what…*

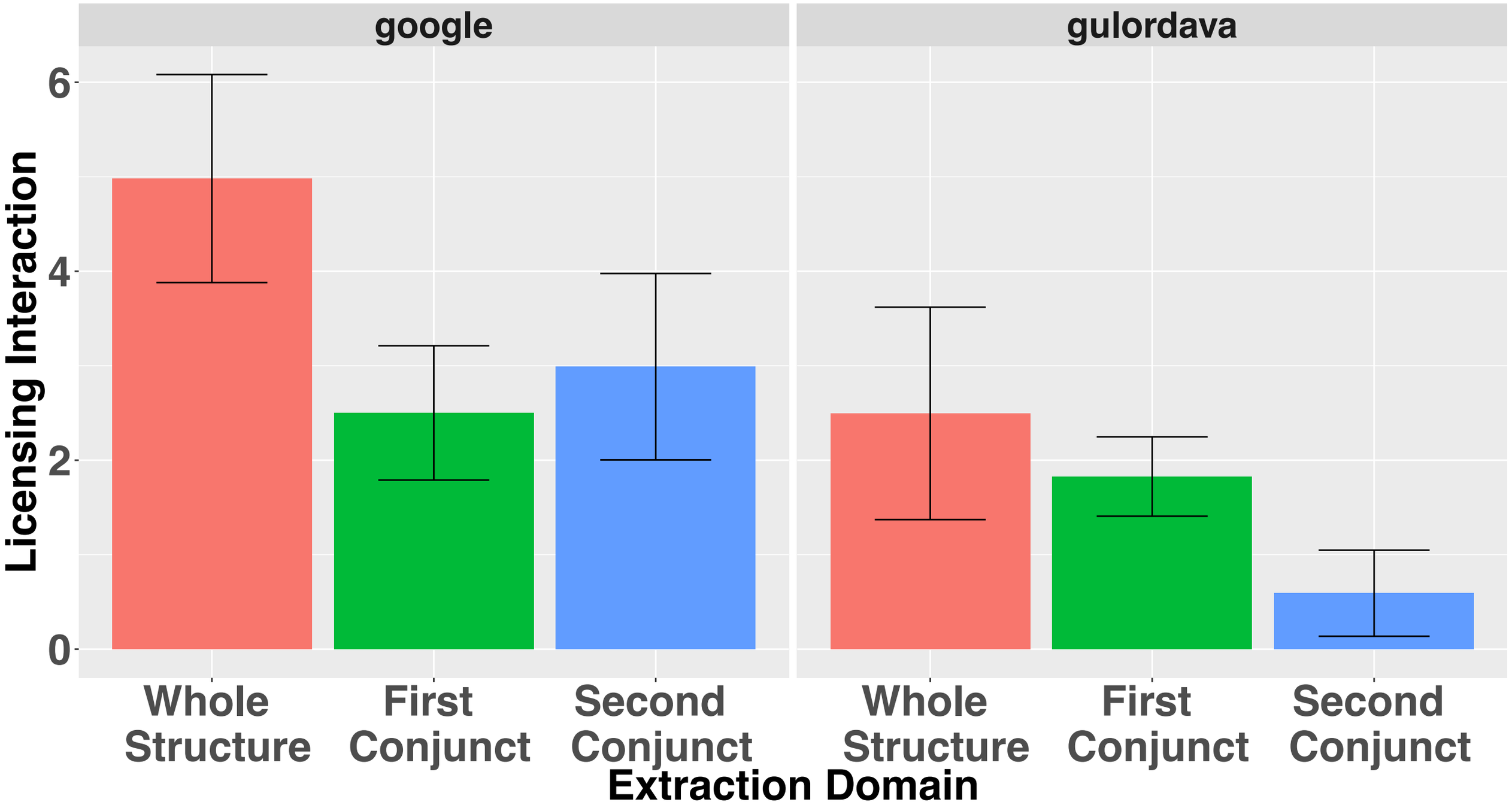
✓ *…you bought ___ at the auction.*
*[Whole Structure]*

* *… you bought ___ and a table at the auction.*
*[First Conjunct]*

* *… you bought a lamp and ___ at the auction.*
*[Second Conjunct]*

# Experiment 12: Coordination Islands

## Significant Reduction for Second Conjunct



# Experiment 13: Left Branch Islands

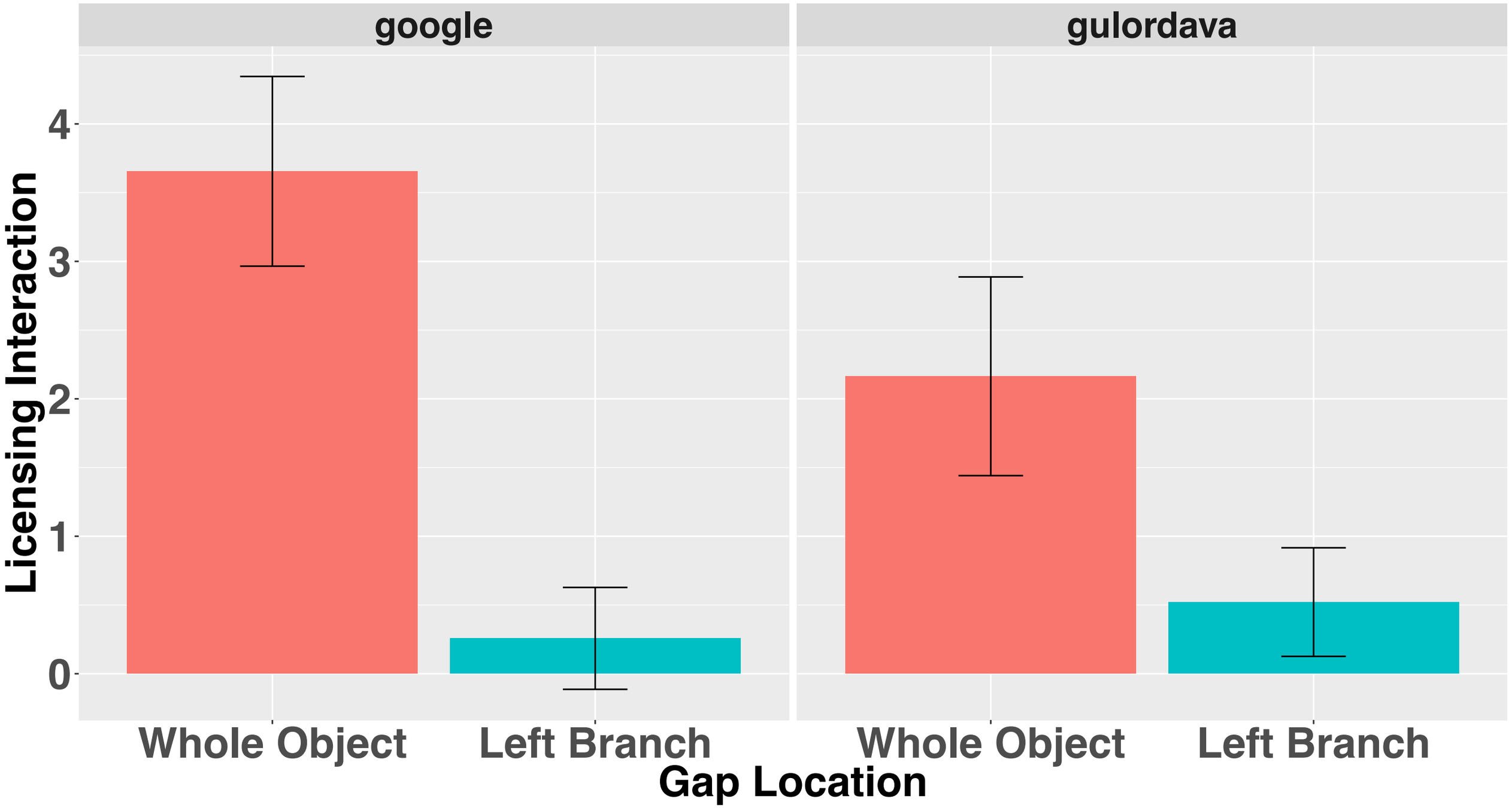
Gaps cannot occur in the **left branch modifiers** under a noun

*We discovered…*

✓ ***…how evil a villain*** *he battled __ yesterday.*
*[Whole Object]*

* ***…how evil*** *he battled __ a villain yesterday.*
*[Left Branch]*

# Experiment 13: Left Branch Islands



# Experiment 14: Sentential Subject Islands

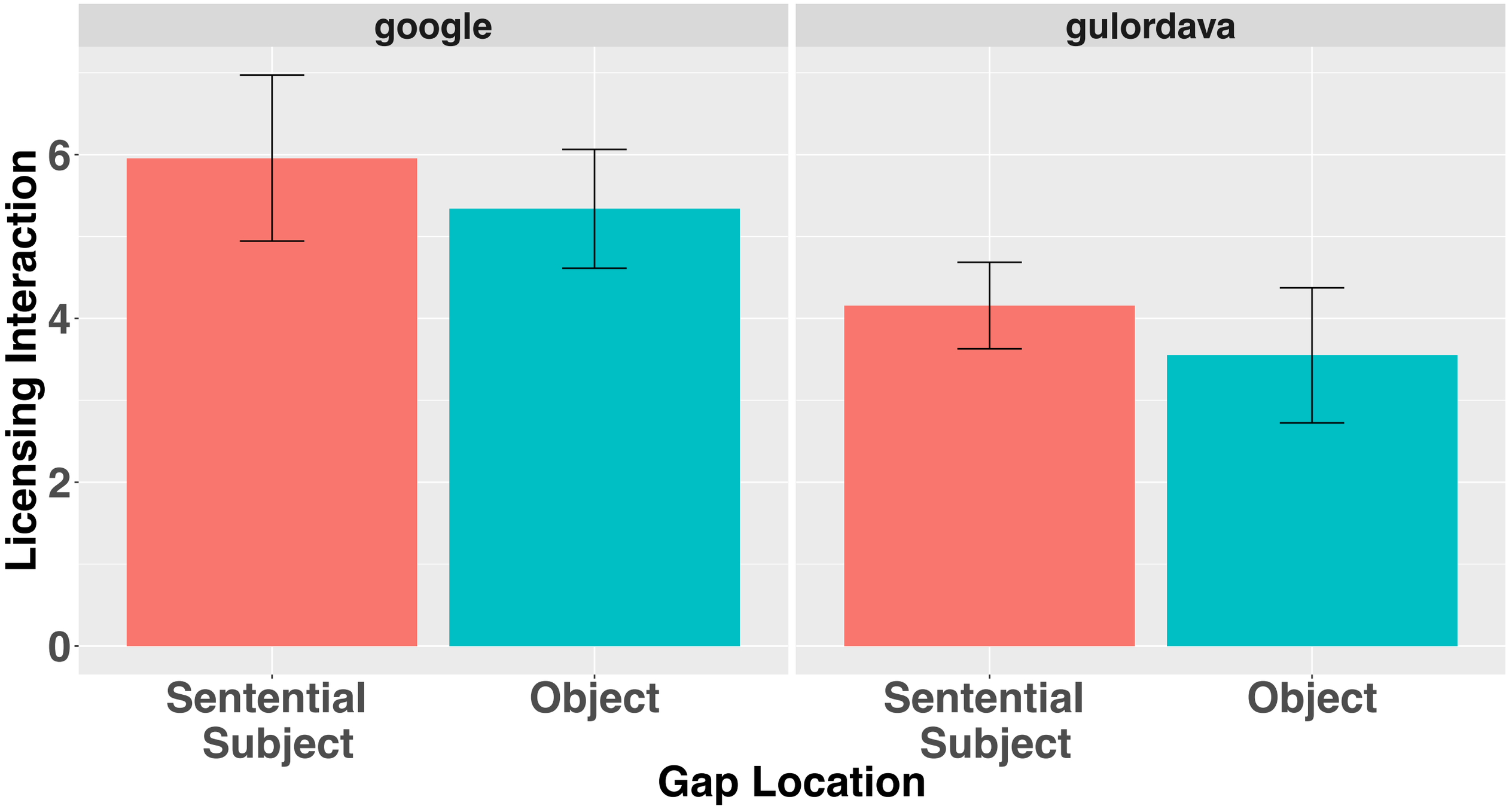
Gaps cannot occur within sentential subjects.

*I know who…*

✓ *…the seniors triumphed over ___ last week.*
*[Object]*

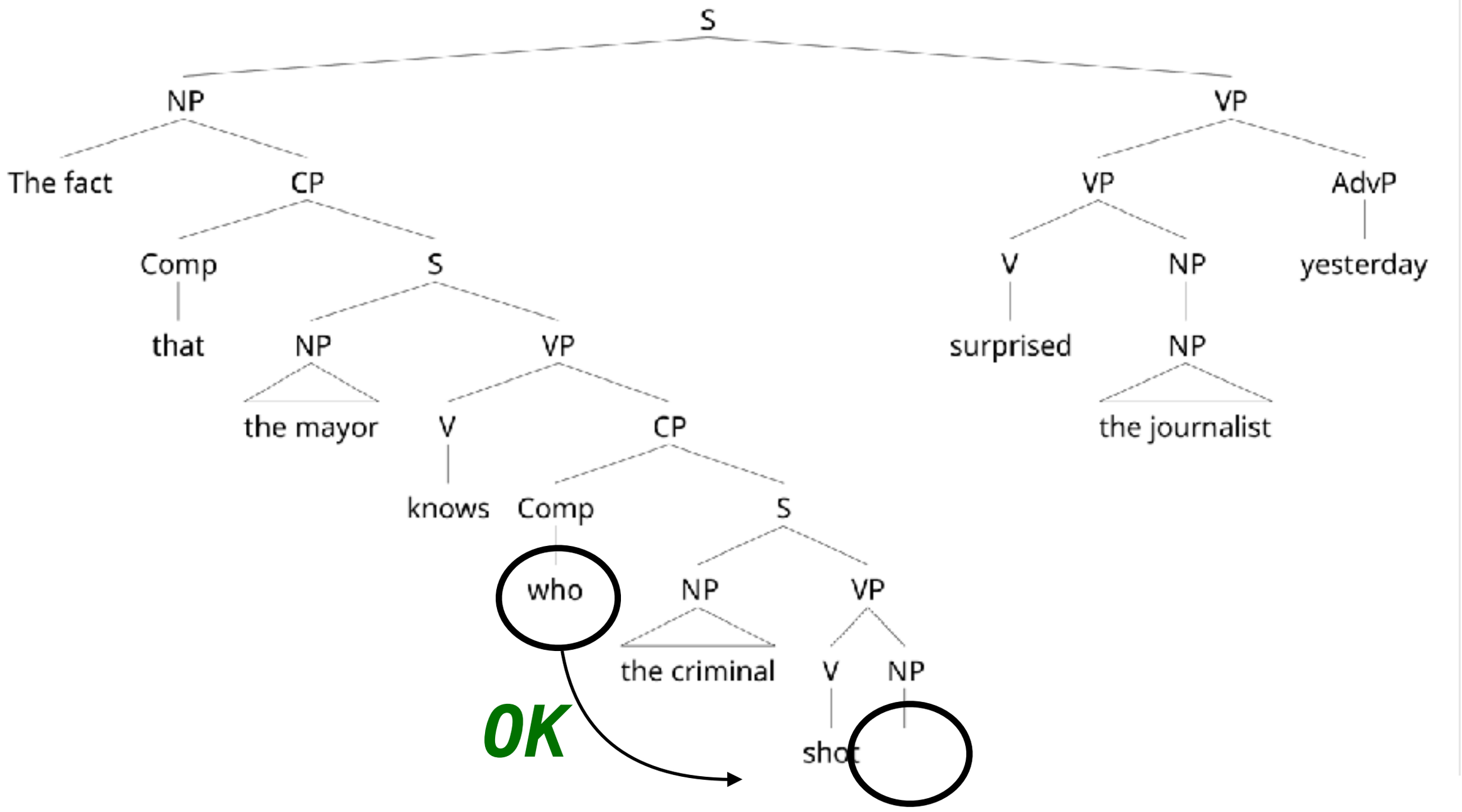
* *…for the seniors to triumph over ___ will be difficult.*
*[Sentential Subject]*

# Experiment 14: Sentential Subject Islands



# Syntactic Hierarchy

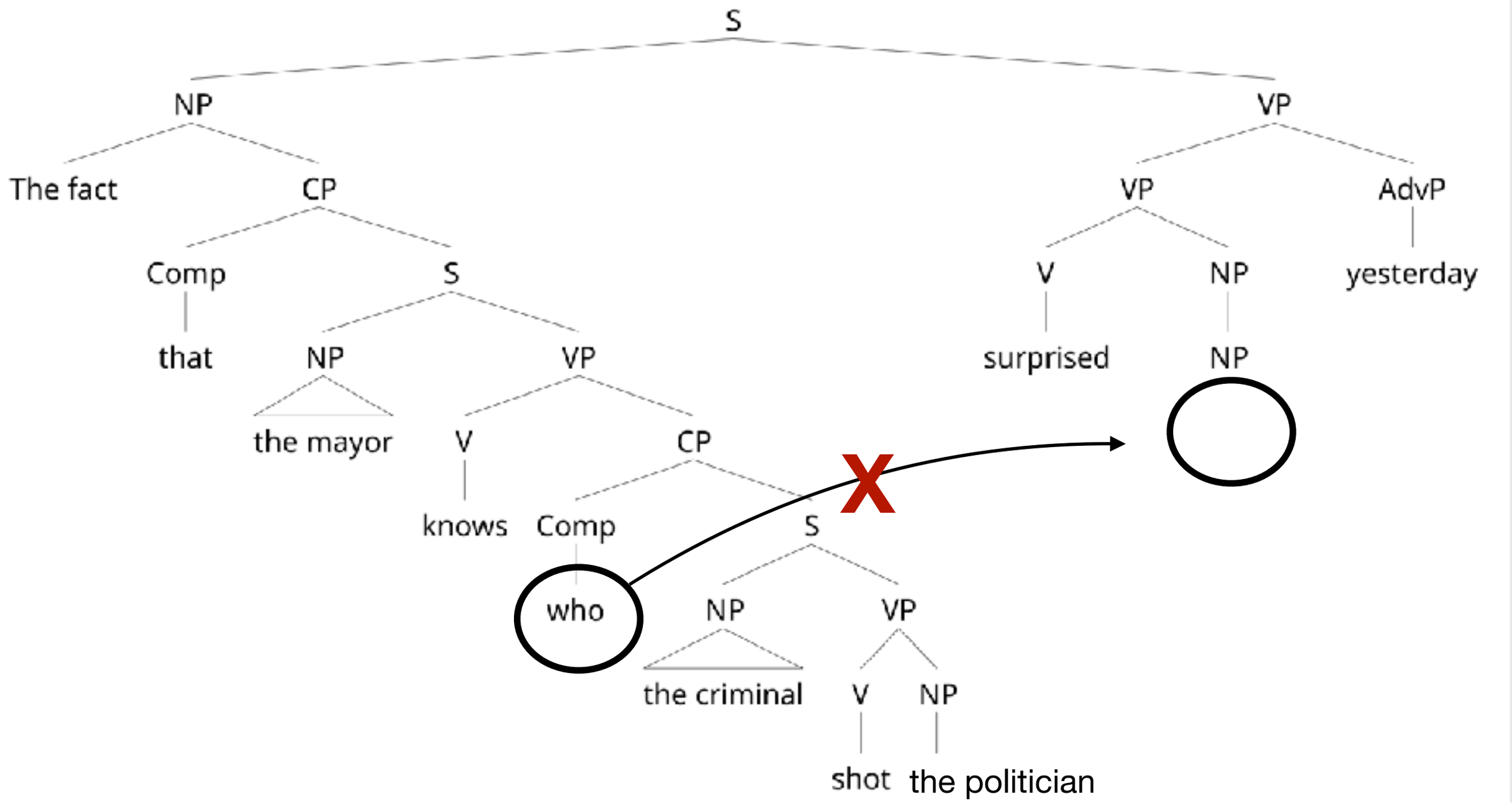
*Filler must **c-command** gap*




✓**The fact that the mayor knows who the criminal shot ___ surprised the journalist yesterday.**
**[Subject Clause]**

# Syntactic Hierarchy

*Filler must **c-command** gap*




∗ **The fact that the mayor knows who the criminal shot the politician surprised ___ yesterday.**
**[Matrix Clause]**

# Syntactic Hierarchy

## Post Gap Material