

# Grammatical Knowledge is Fundamentally Probabilistic

Roger Levy

UC San Diego  
Department of Linguistics  
Annual Meeting of the Linguistic Society of America

10 January 2015

# Introduction

A couple of central questions in linguistics:

# Introduction

A couple of central questions in linguistics:

- ▶ How do we know when relative preferences among strings genuinely reflect grammatical knowledge (“competence”)?

# Introduction

A couple of central questions in linguistics:

- ▶ How do we know when relative preferences among strings genuinely reflect grammatical knowledge (“competence”)?
- ▶ Is the content of grammatical knowledge purely categorical, or does it include a probabilistic component?

# Introduction

- ▶ Desiderata for explaining relative preferences through appeal to grammatical knowledge

# Introduction

- ▶ Desiderata for explaining relative preferences through appeal to grammatical knowledge
  - ▶ The preferences cannot be accounted for through extralinguistic knowledge or independently motivated performance constraints

# Introduction

- ▶ Desiderata for explaining relative preferences through appeal to grammatical knowledge
  - ▶ The preferences cannot be accounted for through extralinguistic knowledge or independently motivated performance constraints
  - ▶ Hypothesizing the grammatical knowledge improves descriptive adequacy

# Introduction

- ▶ Desiderata for explaining relative preferences through appeal to grammatical knowledge
  - ▶ The preferences cannot be accounted for through extralinguistic knowledge or independently motivated performance constraints
  - ▶ Hypothesizing the grammatical knowledge improves descriptive adequacy
- ▶ What types of relative preferences might reflect probabilistic, or gradient, knowledge?

# Introduction

- ▶ Desiderata for explaining relative preferences through appeal to grammatical knowledge
  - ▶ The preferences cannot be accounted for through extralinguistic knowledge or independently motivated performance constraints
  - ▶ Hypothesizing the grammatical knowledge improves descriptive adequacy
- ▶ What types of relative preferences might reflect probabilistic, or gradient, knowledge?
  - ▶ If, among structures that are grammatically licensed, some are “better” than (reliably preferred to) others

# Today's argument

*Does the content of grammatical knowledge include a probabilistic component?*

- ▶ In the domain of syntax, this question remains unresolved

# Today's argument

*Does the content of grammatical knowledge include a probabilistic component?*

- ▶ In the domain of syntax, this question remains unresolved
- ▶ (In phonology, a clear “yes” —Hayes, Albright, Wilson, . . .)

# Today's argument

*Does the content of grammatical knowledge include a probabilistic component?*

- ▶ In the domain of syntax, this question remains unresolved
- ▶ (In phonology, a clear “yes” —Hayes, Albright, Wilson, . . .)
- ▶ Spirited “yes” arguments have been made, but they're not conclusive

# Today's argument

*Does the content of grammatical knowledge include a probabilistic component?*

- ▶ In the domain of syntax, this question remains unresolved
- ▶ (In phonology, a clear “yes” —Hayes, Albright, Wilson, . . .)
- ▶ Spirited “yes” arguments have been made, but they're not conclusive
- ▶ I will adduce a (hopefully) stronger argument

# Argument from acquisition

**Argument:** learning of probabilistic grammars (Horning, 1969) may be more tractable than learning of non-probabilistic grammars (Gold, 1967) (Manning, 2003; Perfors et al., 2011; Clark and Lappin, 2009)

# Argument from acquisition

**Argument:** learning of probabilistic grammars (Horning, 1969) may be more tractable than learning of non-probabilistic grammars (Gold, 1967) (Manning, 2003; Perfors et al., 2011; Clark and Lappin, 2009)

## **Objections:**

- ▶ Maybe learning bias (a.k.a. UG) leaves the learning problem minimally difficult

# Argument from acquisition

**Argument:** learning of probabilistic grammars (Horning, 1969) may be more tractable than learning of non-probabilistic grammars (Gold, 1967) (Manning, 2003; Perfors et al., 2011; Clark and Lappin, 2009)

## Objections:

- ▶ Maybe learning bias (a.k.a. UG) leaves the learning problem minimally difficult
- ▶ Also, non-probabilistic grammar induction is getting better and better (e.g., Clark and Eyraud, 2006)

# Argument from acquisition

**Argument:** learning of probabilistic grammars (Horning, 1969) may be more tractable than learning of non-probabilistic grammars (Gold, 1967) (Manning, 2003; Perfors et al., 2011; Clark and Lappin, 2009)

## Objections:

- ▶ Maybe learning bias (a.k.a. UG) leaves the learning problem minimally difficult
- ▶ Also, non-probabilistic grammar induction is getting better and better (e.g., Clark and Eyraud, 2006)
- ▶ Or, probabilities may be used during acquisition and then discarded in encoding a categorical competence grammar

# Syntactic attachment preference

**Argument:** there are systematic attachment preferences for configurations such as

[VP V [VP V ...] ...] Adv

e.g.,

*John said that Mary left yesterday*

and those preferences match corpus frequencies (Jurafsky, 1996, 2003; Levy, 2008; MacDonald and Thornton, 2009)

# Syntactic attachment preference

**Argument:** there are systematic attachment preferences for configurations such as

[VP V [VP V ...] ...] Adv

e.g.,

*John said that Mary left yesterday*

and those preferences match corpus frequencies (Jurafsky, 1996, 2003; Levy, 2008; MacDonald and Thornton, 2009)

**Objection:** This type of attachment preference may be due to linear-distance locality preferences (Hawkins, 1994; Gibson, 1998), not structural frequencies

# Lexeme-specific syntactic preference

**Argument:** For a given set of thematic roles, different verbs seem to have idiosyncratically different preferences for syntactic realization (Manning, 2003; Bresnan et al., 2007):

## The dative alternation

[V NP<sub><Goal></sub> NP<sub><Theme></sub>]-preferring

*tell*

*teach*

*charge*

*show*

[V NP<sub><Theme></sub> PP<sub><Goal></sub>]-preferring

*loan*

*bring*

*sell*

*take*

# Lexeme-specific syntactic preference

**Argument:** For a given set of thematic roles, different verbs seem to have idiosyncratically different preferences for syntactic realization (Manning, 2003; Bresnan et al., 2007):

## The dative alternation

| [V NP <sub>&lt;Goal&gt;</sub> NP <sub>&lt;Theme&gt;]-preferring</sub> | [V NP <sub>&lt;Theme&gt;</sub> PP <sub>&lt;Goal&gt;]-preferring</sub> |
|---|---|
| <i>tell</i>   | <i>loan</i>   |
| <i>teach</i>  | <i>bring</i>  |
| <i>charge</i>   | <i>sell</i>   |
| <i>show</i>   | <i>take</i>   |

**Objection:** Lexeme-specific preferences may be epiphenomena of differences in *shades of meaning* among syntactic realizations

# The internal structure of coordinate categories

- ▶ Principle of *Conjoin Likes* (Chomsky, 1965)

$X \rightarrow X \text{ Conj } X$

# The internal structure of coordinate categories

- ▶ Principle of *Conjoin Likes* (Chomsky, 1965)

$X \rightarrow X \text{ Conj } X$

- ▶ Empirically false for gross syntactic category (Sag et al., 1985):

*Pat is a Republican and proud of it (NP and AdjP)*

# The internal structure of coordinate categories

- ▶ Principle of *Conjoin Likes* (Chomsky, 1965)

$X \rightarrow X \text{ Conj } X$

- ▶ Empirically false for gross syntactic category (Sag et al., 1985):

*Pat is a Republican and proud of it (NP and AdjP)*

# The internal structure of coordinate categories

- ▶ Principle of *Conjoin Likes* (Chomsky, 1965)

$X \rightarrow X \text{ Conj } X$

- ▶ Empirically false for gross syntactic category (Sag et al., 1985):

*Pat is a Republican and proud of it (NP and AdjP)*

*That was a rude remark and in very bad taste (NP and PP)*

# The internal structure of coordinate categories

- ▶ Principle of *Conjoin Likes* (Chomsky, 1965)

$X \rightarrow X \text{ Conj } X$

- ▶ Empirically false for gross syntactic category (Sag et al., 1985):

*Pat is a Republican and proud of it (NP and AdjP)*

*That was a rude remark and in very bad taste (NP and PP)*

- ▶ So... was “Conjoin Likes” just wrong?

# The internal structure of coordinate categories

- ▶ Principle of *Conjoin Likes* (Chomsky, 1965)

$$X \rightarrow X \text{ Conj } X$$

- ▶ Empirically false for gross syntactic category (Sag et al., 1985):

*Pat is a Republican and proud of it (NP and AdjP)*

*That was a rude remark and in very bad taste (NP and PP)*

- ▶ So... was “Conjoin Likes” just wrong?
- ▶ Is there anything left for grammar to say about a “tendency” for coordinated categories to be like one another?

- ▶ Unlike-category coordinations *are* easy to find in corpora  
*His son had been friendly, a big fellow of fifty or more, a fishing-boat captain and powerful like the sea*

(Parsed Brown corpus)

# Corpus data

- ▶ Unlike-category coordinations *are* easy to find in corpora

*His son had been friendly, a big fellow of fifty or more, a fishing-boat captain and powerful like the sea*

- ▶ But there is a huge *quantitative* tendency for coordination to be of like categories in corpora

|           |      | Right-hand conjunct |      |
|-----------|------|---------------------|------|
|           |      | NP                  | AdjP |
| Left-hand | NP   | 1308                | 8    |
| Conjunct  | AdjP | 6                   | 114  |

(Parsed Brown corpus)

## *Conjoin Likes* as a gradient grammatical constraint?

- ▶ Perhaps this is gradient *Conjoin Likes*!

## *Conjoin Likes* as a gradient grammatical constraint?

- ▶ Perhaps this is gradient *Conjoin Likes*!
- ▶ But should we really attribute this to the grammar proper?

## *Conjoin Likes* as a gradient grammatical constraint?

- ▶ Perhaps this is gradient *Conjoin Likes*!
- ▶ But should we really attribute this to the grammar proper?
- ▶ We need:

## *Conjoin Likes* as a gradient grammatical constraint?

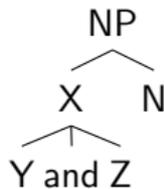
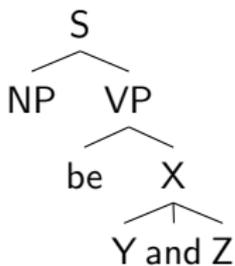
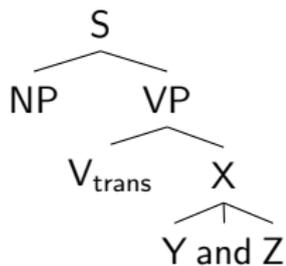
- ▶ Perhaps this is gradient *Conjoin Likes*!
- ▶ But should we really attribute this to the grammar proper?
- ▶ We need:
  - ▶ A precise proposal of gradient *Conjoin Likes*

# *Conjoin Likes* as a gradient grammatical constraint?

- ▶ Perhaps this is gradient *Conjoin Likes*!
- ▶ But should we really attribute this to the grammar proper?
- ▶ We need:
  - ▶ A precise proposal of gradient *Conjoin Likes*
  - ▶ Evidence of relative preference for like-conjunct coordinations that **controls for meaning and external context**

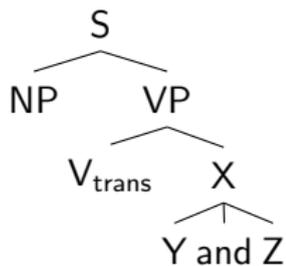
# Conjoin Likes as a gradient grammatical constraint?

A cautionary thought experiment:

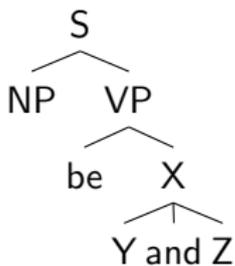


# Conjoin Likes as a gradient grammatical constraint?

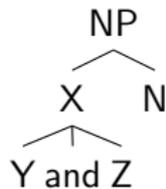
A cautionary thought experiment:



↓  
NP and NP



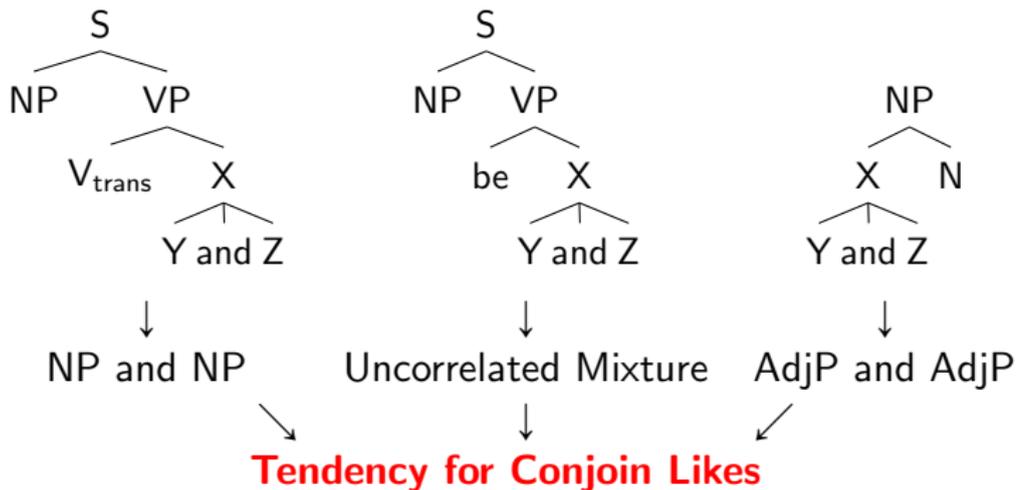
↓  
Uncorrelated Mixture



↓  
AdjP and AdjP

# Conjoin Likes as a gradient grammatical constraint?

A cautionary thought experiment:



# Model

What “gradient coordination of like categories” means:

$$P(F_1, F_2 | M_1, M_2)$$

is especially high when  $F_1$  and  $F_2$  are “like” in the traditional sense of

$X \rightarrow X$  and  $X$

Fully technically:

$$\text{pMI}(F_1, F_2 | M_1, M_2) = \log \frac{P(F_1, F_2 | M_1, M_2)}{P(F_1 | M_1)P(F_2 | M_2)}$$

is monotonically increasing in the structural similarity of  $F_1$  and  $F_2$

# Empirical prediction

If more probable forms are gradually “more grammatical” to the native speaker. . .

# Empirical prediction

If more probable forms are gradiently “more grammatical” to the native speaker. . . then like-category coordinations should be judged more natural, or acceptable, than unlike-category coordinations

# Empirical prediction

If more probable forms are gradiently “more grammatical” to the native speaker. . . then like-category coordinations should be judged more natural, or acceptable, than unlike-category coordinations

More precisely, if

$$\left\{ \begin{array}{l} \text{pMI}(\text{AdjP}, \text{AdjP} | M_1, M_2) \\ \text{pMI}(\text{NP}, \text{NP} | M_1, M_2) \end{array} \right\} > \left\{ \begin{array}{l} \text{pMI}(\text{AdjP}, \text{NP} | M_1, M_2) \\ \text{pMI}(\text{NP}, \text{AdjP} | M_1, M_2) \end{array} \right\}$$

then

$$\underbrace{\log P(\text{AdjP}, \text{AdjP} | M_1, M_2) + \log P(\text{NP}, \text{NP} | M_1, M_2)}_{\vee}$$

$$\underbrace{\log P(\text{AdjP}, \text{NP} | M_1, M_2) + \log P(\text{NP}, \text{AdjP} | M_1, M_2)}$$

and like-conjunct coordinations should be superadditively good.

# Experiment 1

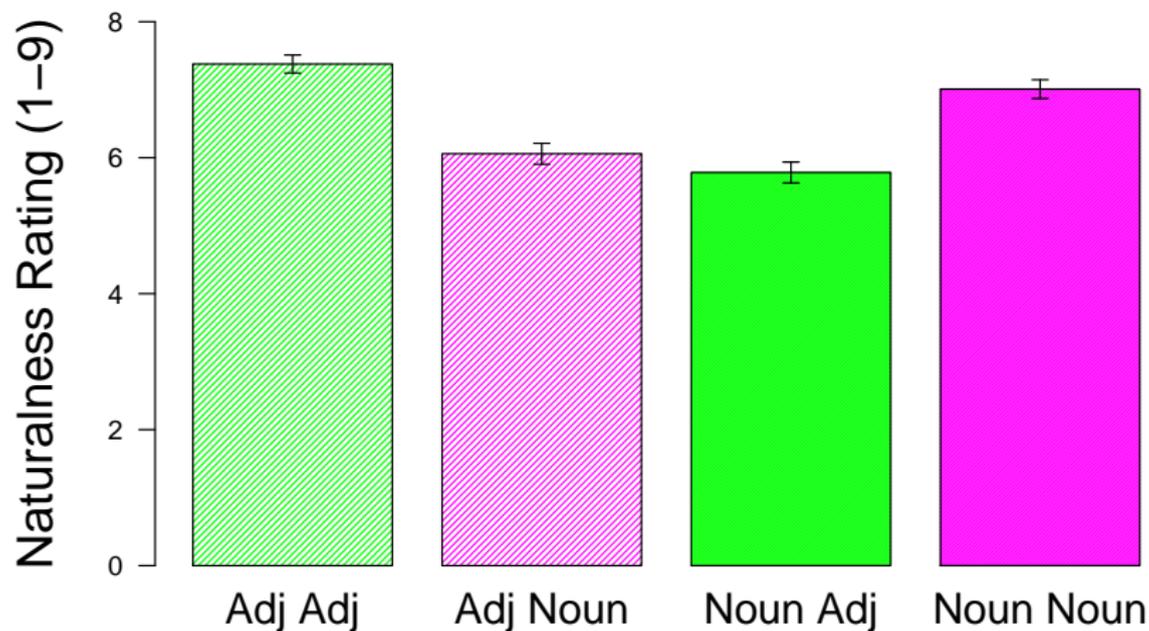
Acceptability judgment study (scale of 1–9):

|   |             |
|---|-------------|
| <i>Pat is a Republican and a freak.</i> | [Noun Noun] |
| <i>Pat is a Republican and freaky.</i>  | [Noun Adj ] |
| <i>Pat is Republican and a freak.</i>   | [Adj Noun]  |
| <i>Pat is Republican and freaky.</i>    | [Adj Adj ]  |

(Baseline: *The children decorated the sparkling ornaments onto the tree was a 4.*)

(All experiments: Mechanical Turk native speakers, 12 items, the same 21 fillers (all “good”),  $\geq 75$  subjects.)

# Experiment 1: Results



*The gradient preference for coordination of unlike categories is pretty strong!*

# Greater explanatory power of gradient constraints

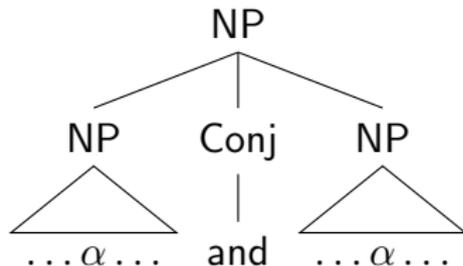
- ▶ We saw that “Conjoin Likes” is categorically false, but “probabilistically” true

# Greater explanatory power of gradient constraints

- ▶ We saw that “Conjoin Likes” is categorically false, but “probabilistically” true
- ▶ But why stop at major syntactic categories—what about category-*internal* structure (Johnson, 1998; Klein and Manning, 2003)?

# Greater explanatory power of gradient constraints

- ▶ We saw that “Conjoin Likes” is categorically false, but “probabilistically” true
- ▶ But why stop at major syntactic categories—what about category-*internal* structure (Johnson, 1998; Klein and Manning, 2003)?
- ▶ Such a grammatical preference has previously been explored under the rubric of PARALLELISM (Frazier et al., 1984; Hale et al., 2006; Dubey et al., 2008)



# NP-internal parallelism: the genitive alternation

Postnominal

The future of our country

The base of the lamp

The tail of a cat

~

~

~

Prenominal

Our country's future

The lamp's base

A cat's tail

# NP-internal parallelism: the genitive alternation

Postnominal

The future of our country

The base of the lamp

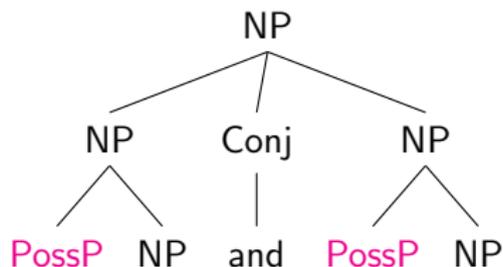
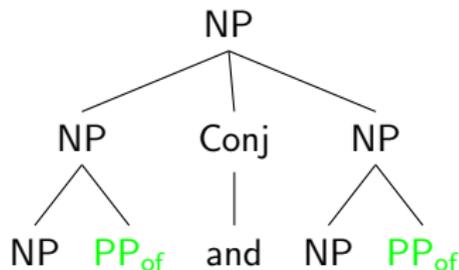
The tail of a cat

Prenominal

Our country's future

The lamp's base

A cat's tail



# NP-internal parallelism: the genitive alternation

Postnominal

The future of our country

The base of the lamp

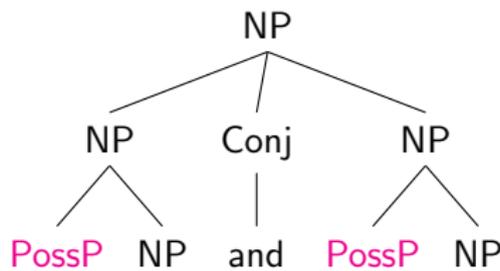
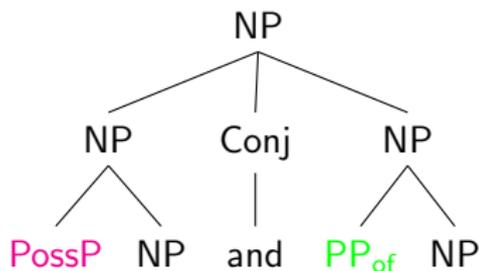
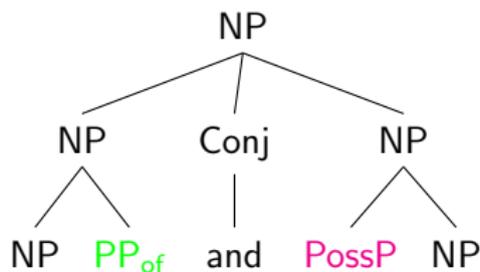
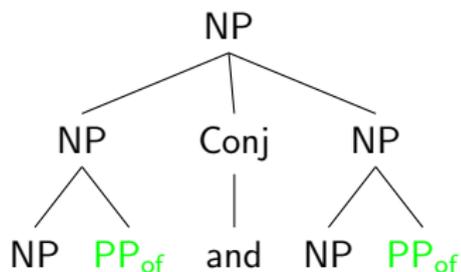
The tail of a cat

Prenominal

~ Our country's future

~ The lamp's base

~ A cat's tail



# Corpus data on genitive alternation parallelism

|               |      | Right Conjunct |     |
|---------------|------|----------------|-----|
|               |      | Post           | Pre |
| Left Conjunct | Post | 77             | 15  |
|               | Pre  | 20             | 39  |

- ▶ There is also strong evidence for a parallelism preference in the genitive alternation. . .

# Corpus data on genitive alternation parallelism

|               |      | Right Conjunct |     |
|---------------|------|----------------|-----|
|               |      | Post           | Pre |
| Left Conjunct | Post | 77             | 15  |
|               | Pre  | 20             | 39  |

- ▶ There is also strong evidence for a parallelism preference in the genitive alternation. . .
- ▶ . . .but once again this analysis fails to control for external context and conjunct meanings  $M_1, M_2$

# Corpus data on genitive alternation parallelism

|               |      | Right Conjunct |     |
|---------------|------|----------------|-----|
|               |      | Post           | Pre |
| Left Conjunct | Post | 77             | 15  |
|               | Pre  | 20             | 39  |

- ▶ There is also strong evidence for a parallelism preference in the genitive alternation. . .
- ▶ . . .but once again this analysis fails to control for external context and conjunct meanings  $M_1, M_2$
- ▶ We can control this more tightly with an experiment

## Experiment 2: genitive parallelism

Acceptability judgment study (scale of 1–9):

*Terry assembled...*

*... the frame of the chair and the base of the lamp.* [Post Post]

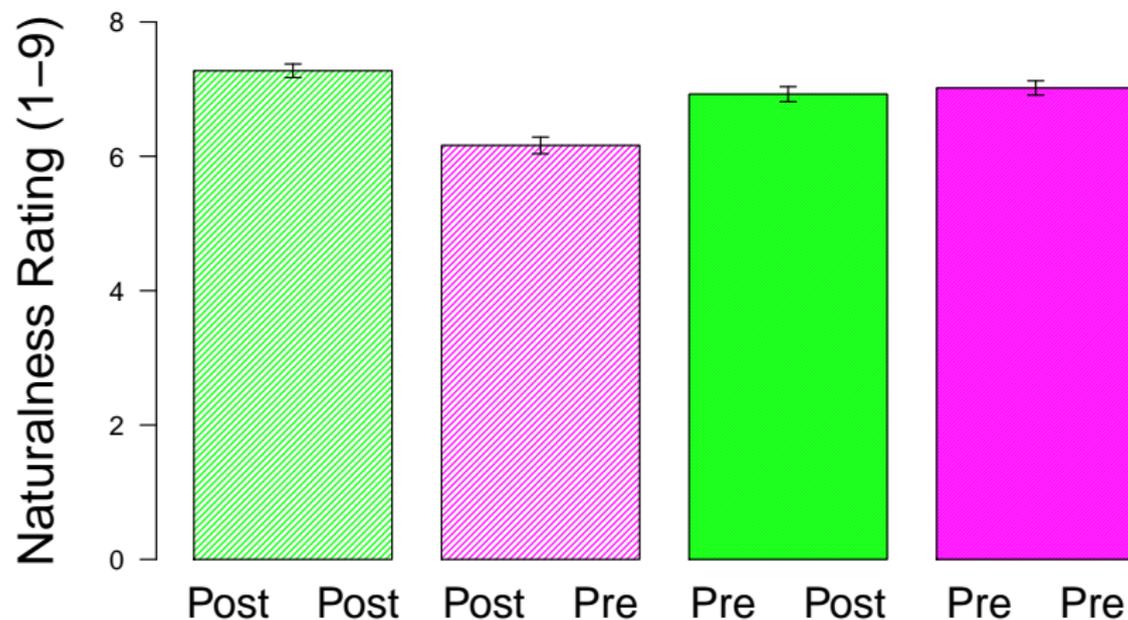
*... the frame of the chair and the lamp's base.* [Post Pre ]

*... the chair's frame and the base of the lamp.* [Pre Post]

*... the chair's frame and the lamp's base.* [Pre Pre ]

(Baseline: *The children decorated the sparkling ornaments onto the tree was a*  
4.)

## Experiment 2: Results

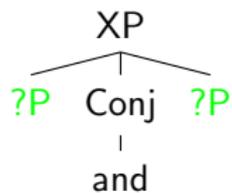


*There is also a preference for parallelism among realizations of the genitive alternation!*

# Searching for deeper parallelism effects

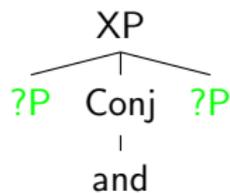
# Searching for deeper parallelism effects

Like categories!

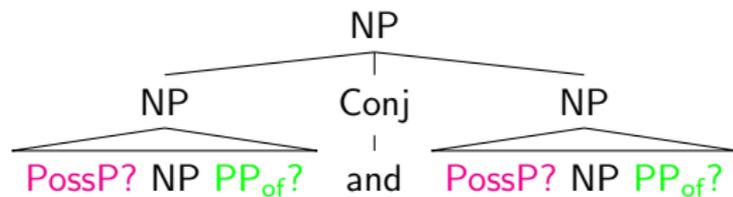


# Searching for deeper parallelism effects

Like categories!

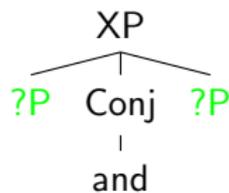


Genitive parallelism!

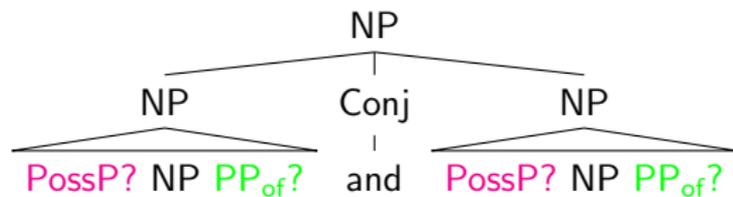


# Searching for deeper parallelism effects

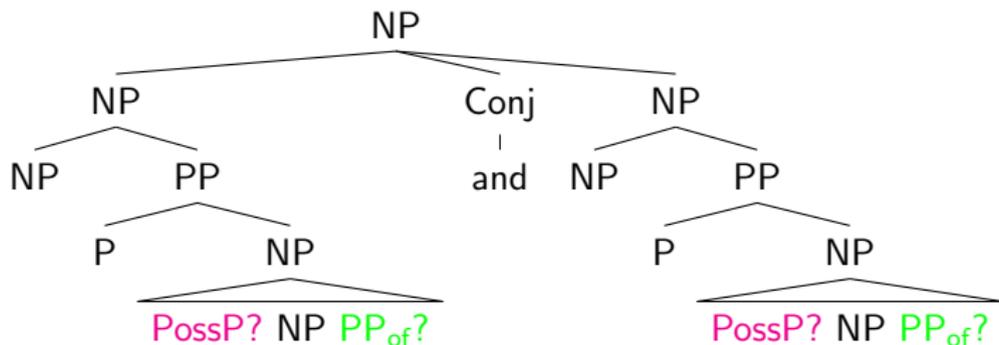
Like categories!



Genitive parallelism!



**Embedded-genitive parallelism???**



## Experiment 3: parallelism of embedded genitive

*Kim is full of...*

*...skepticism about the future of our country and pessimism regarding the prospects of our children.*

[Post Post]

*...skepticism about the future of our country and pessimism regarding our children's prospects.*

[Post Pre ]

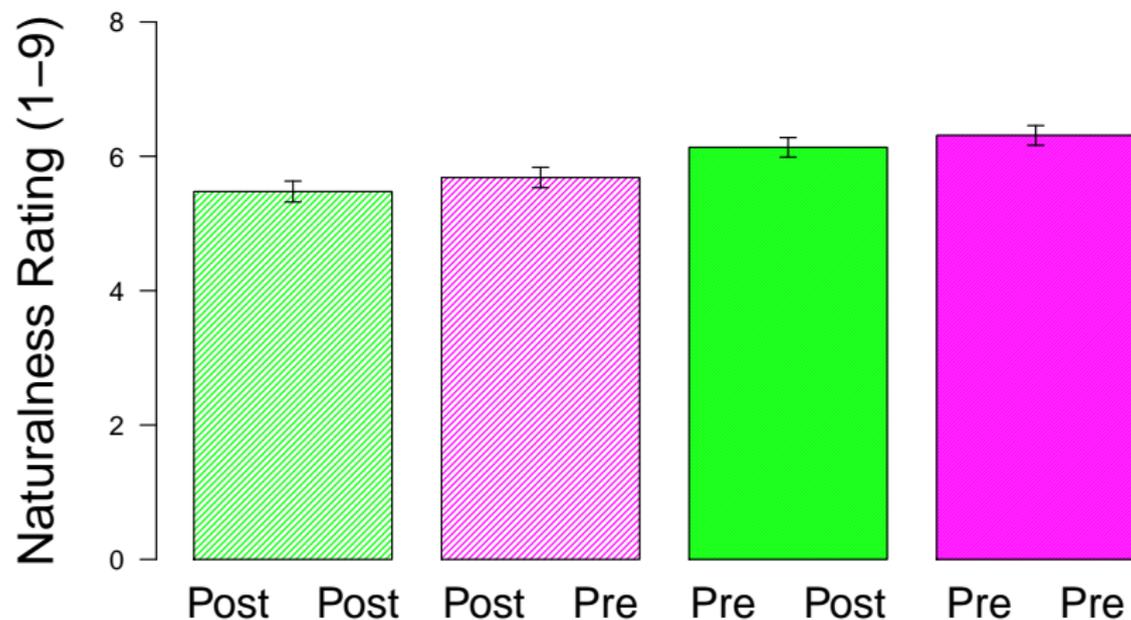
*...skepticism about our country's future and pessimism regarding the prospects of our children.*

[Pre Post]

*...skepticism about our country's future and pessimism regarding our children's prospects.*

[Pre Pre ]

## Experiment 3: Results



*This deep in a coordinate NP structure, the parallelism preference has disappeared!*

# Ruling out priming

**Argument thus far:**

# Ruling out priming

## Argument thus far:

- ▶ Categorical *Conjoin Likes* does not exist

# Ruling out priming

## Argument thus far:

- ▶ Categorical *Conjoin Likes* does not exist
- ▶ But gradient *Conjoin Likes* does

# Ruling out priming

## Argument thus far:

- ▶ Categorical *Conjoin Likes* does not exist
- ▶ But gradient *Conjoin Likes* does
- ▶ We've ruled out external context and conjunct meaning as confounds

# Ruling out priming

## Argument thus far:

- ▶ Categorical *Conjoin Likes* does not exist
- ▶ But gradient *Conjoin Likes* does
- ▶ We've ruled out external context and conjunct meaning as confounds
- ▶ *Conjoin Likes* constitutes non-categorical, probabilistic grammatical knowledge

# Ruling out priming

## Argument thus far:

- ▶ Categorical *Conjoin Likes* does not exist
- ▶ But gradient *Conjoin Likes* does
- ▶ We've ruled out external context and conjunct meaning as confounds
- ▶ *Conjoin Likes* constitutes non-categorical, probabilistic grammatical knowledge

# Ruling out priming

## Argument thus far:

- ▶ Categorical *Conjoin Likes* does not exist
- ▶ But gradient *Conjoin Likes* does
- ▶ We've ruled out external context and conjunct meaning as confounds
- ▶ *Conjoin Likes* constitutes non-categorical, probabilistic grammatical knowledge

**Possible objection:** the gradient *Conjoin Likes* preference has another, independently motivated explanation

# Ruling out priming

## Argument thus far:

- ▶ Categorical *Conjoin Likes* does not exist
- ▶ But gradient *Conjoin Likes* does
- ▶ We've ruled out external context and conjunct meaning as confounds
- ▶ *Conjoin Likes* constitutes non-categorical, probabilistic grammatical knowledge

**Possible objection:** the gradient *Conjoin Likes* preference has another, independently motivated explanation

**Proposed gradient knowledge**

**Independent explanation**

# Ruling out priming

## Argument thus far:

- ▶ Categorical *Conjoin Likes* does not exist
- ▶ But gradient *Conjoin Likes* does
- ▶ We've ruled out external context and conjunct meaning as confounds
- ▶ *Conjoin Likes* constitutes non-categorical, probabilistic grammatical knowledge

**Possible objection:** the gradient *Conjoin Likes* preference has another, independently motivated explanation

### Proposed gradient knowledge

Local attachment preference  
(*said... left... tomorrow*)

### Independent explanation

Memory-driven locality

# Ruling out priming

## Argument thus far:

- ▶ Categorical *Conjoin Likes* does not exist
- ▶ But gradient *Conjoin Likes* does
- ▶ We've ruled out external context and conjunct meaning as confounds
- ▶ *Conjoin Likes* constitutes non-categorical, probabilistic grammatical knowledge

**Possible objection:** the gradient *Conjoin Likes* preference has another, independently motivated explanation

### Proposed gradient knowledge

Local attachment preference  
(*said... left... tomorrow*)

Verb-specific argument realization preferences

Gradient *Conjoin Likes*

### Independent explanation

Memory-driven locality

Constructional meaning differences

**Syntactic priming?**

## Experiment 4: parallelism without coordination

Subject/object configuration rather than coordinated object NPs:

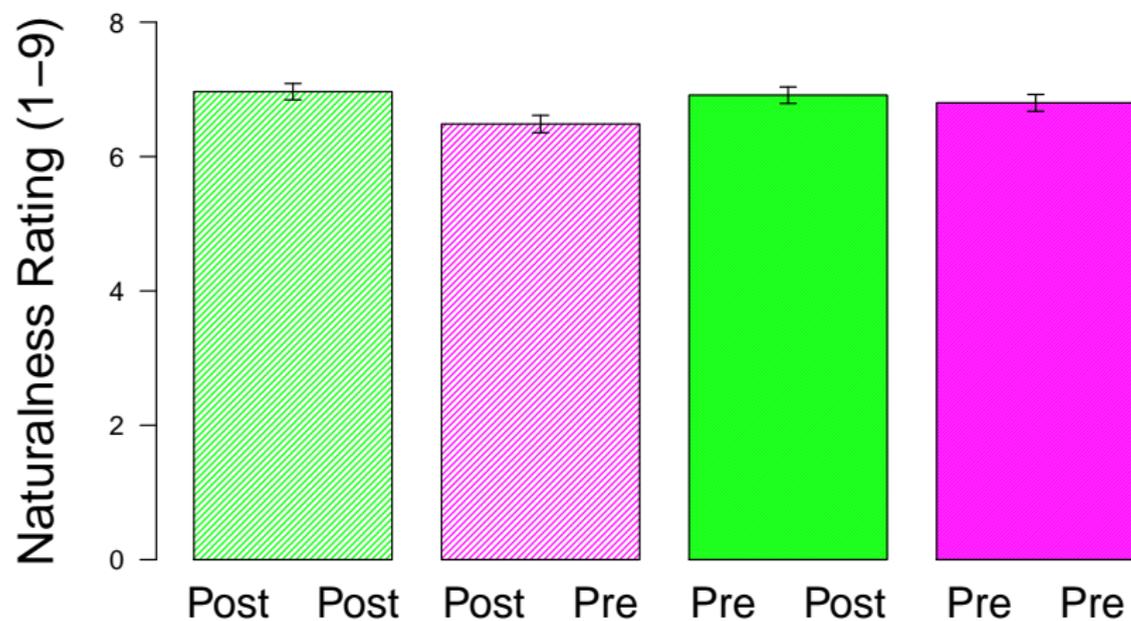
The future of our country will determine the prospects of our children. [Post Post]

The future of our country will determine our children's prospects. [Post Pre ]

Our country's future will determine the prospects of our children. [Pre Post]

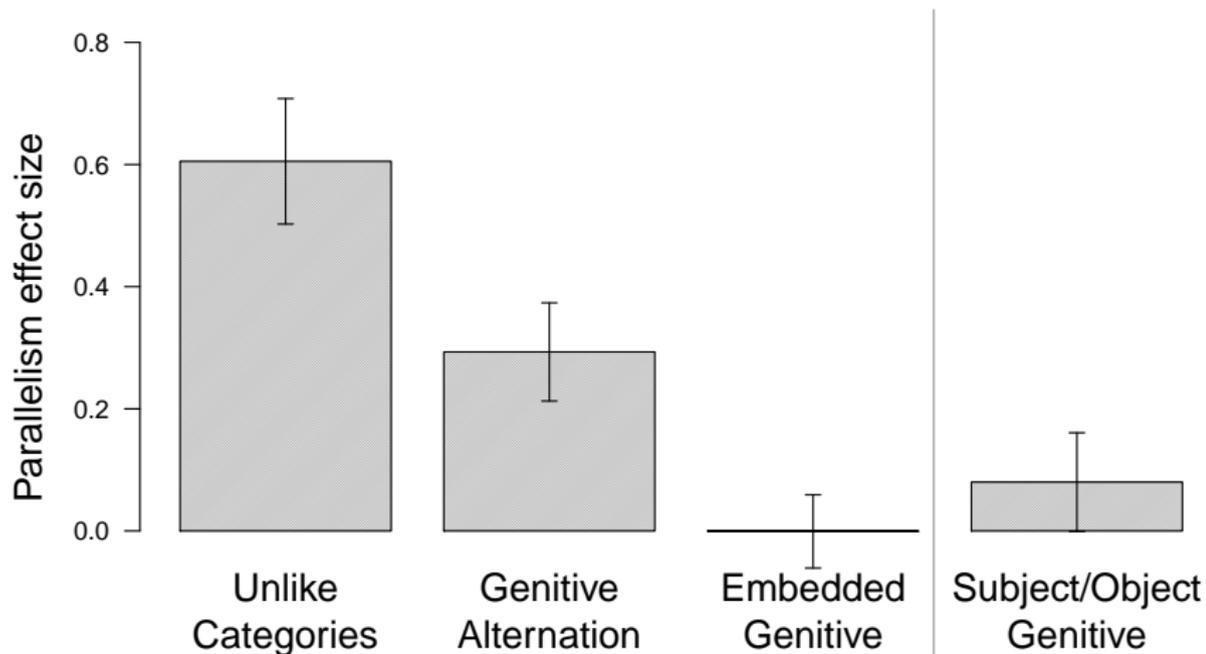
Our country's future will determine our children's prospects. [Pre Pre ]

## Experiment 4: Results

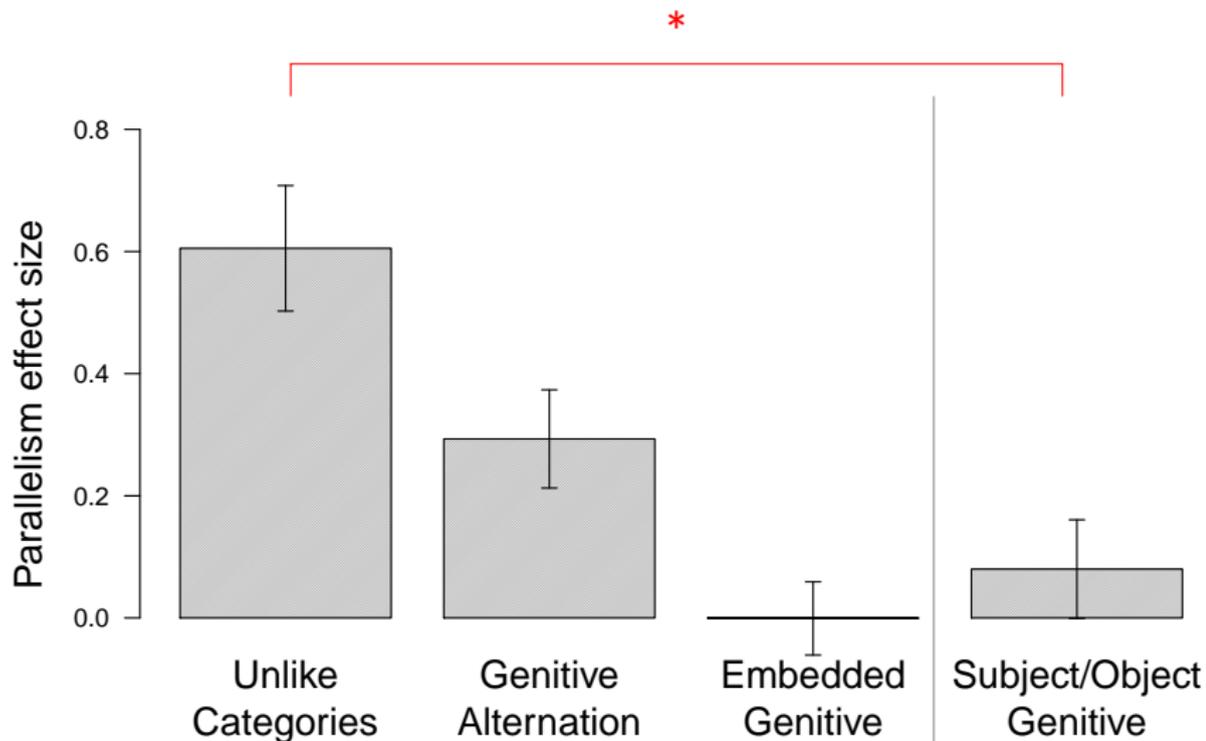


*There is a parallelism effect here! So...?*

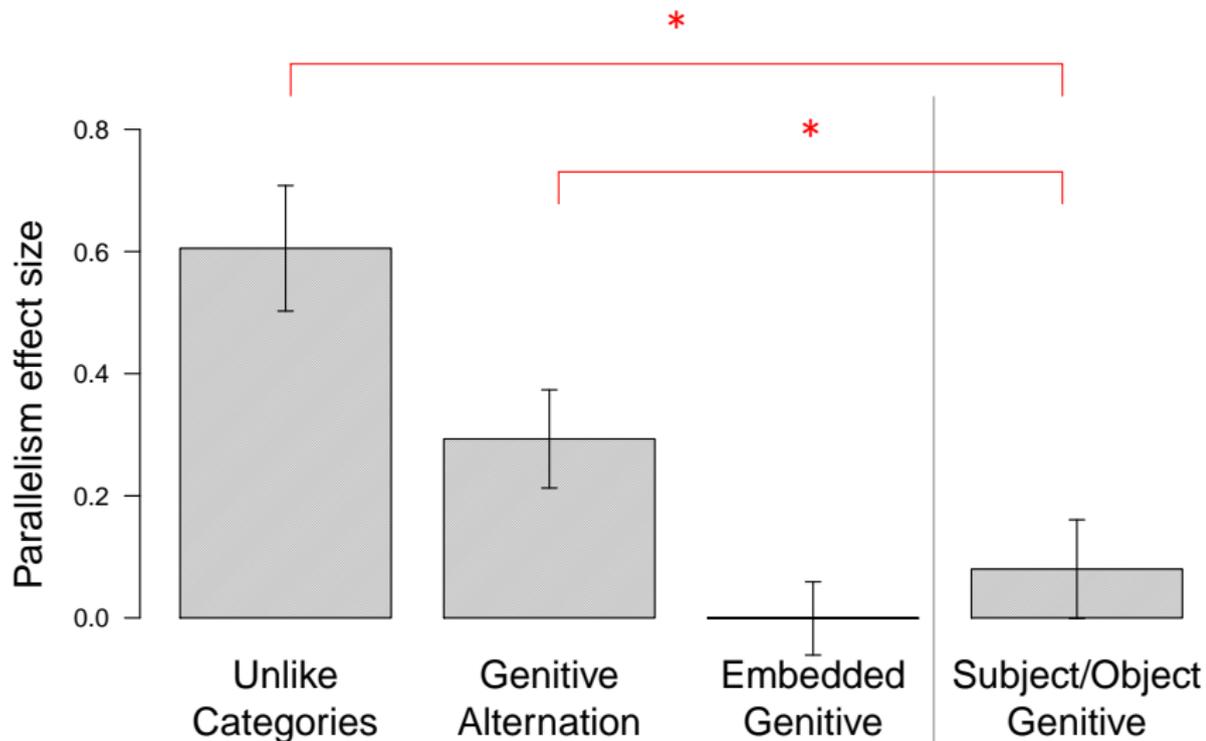
# Strength of parallelism across 4 experiments



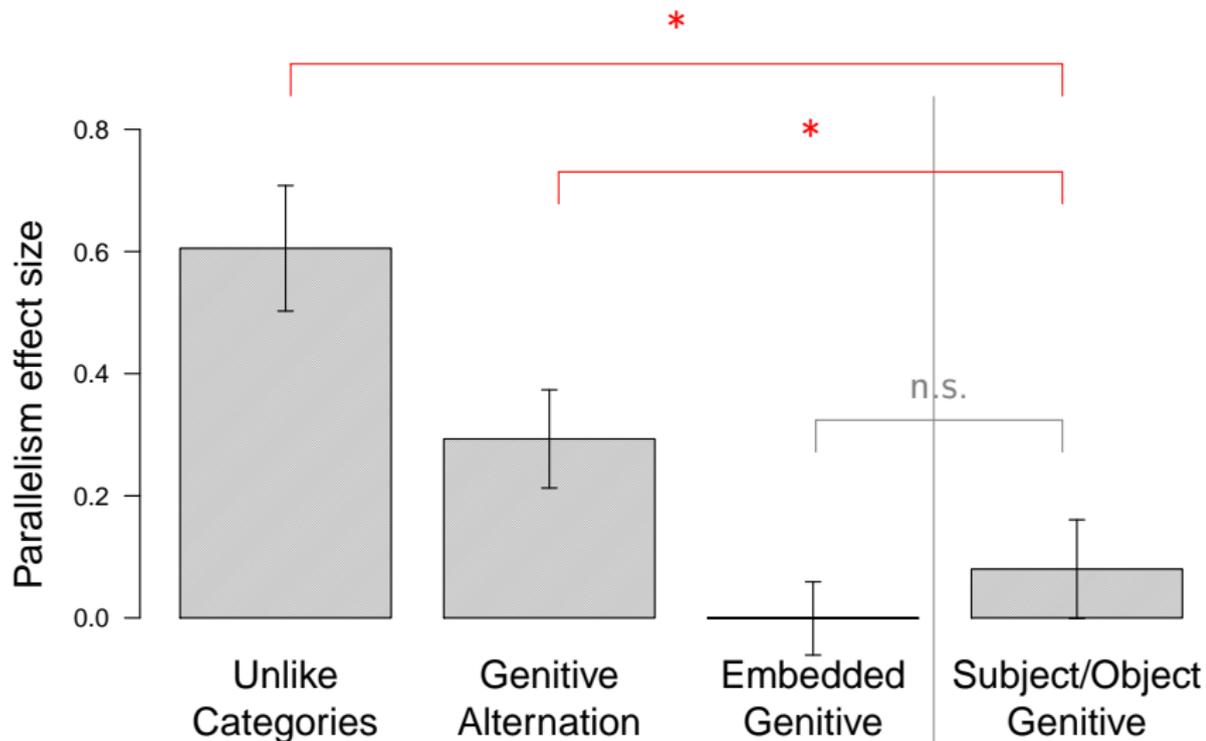
# Strength of parallelism across 4 experiments



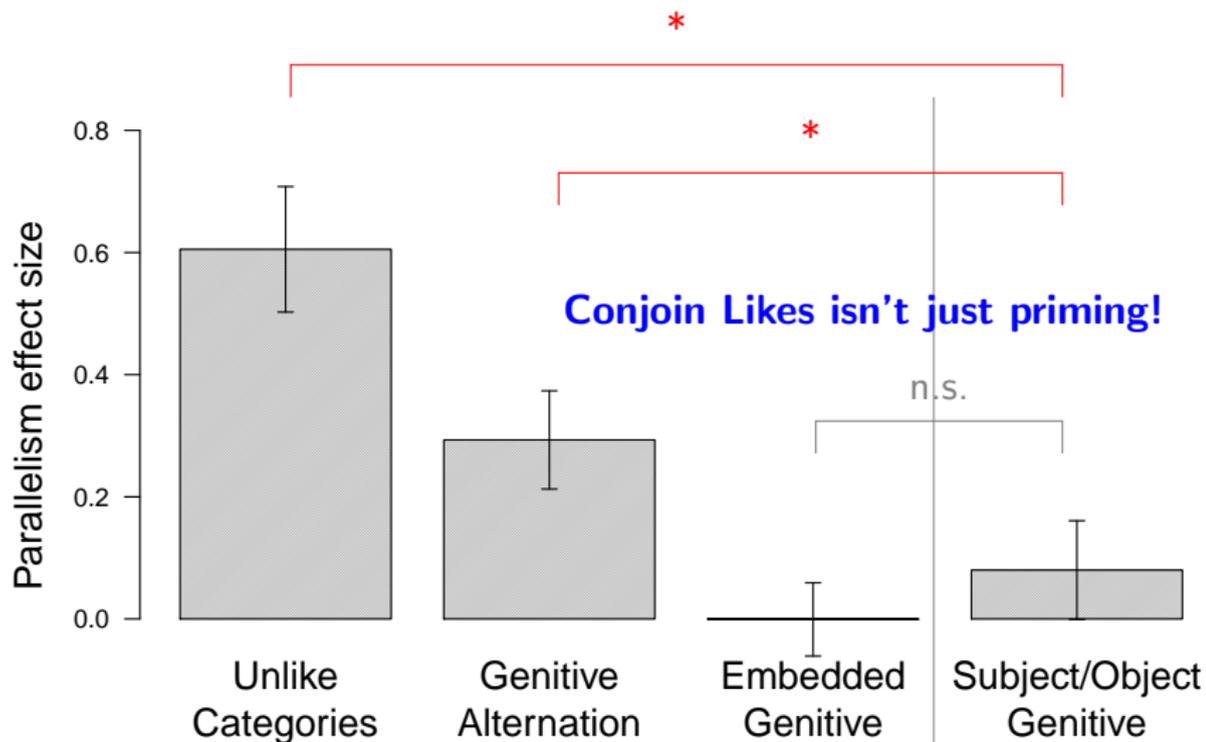
# Strength of parallelism across 4 experiments



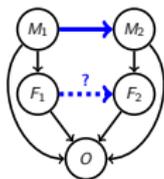
# Strength of parallelism across 4 experiments



# Strength of parallelism across 4 experiments

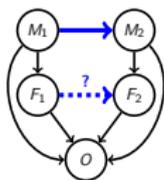


# Discussion



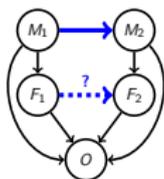
- ▶ Grammar has very little to say about **categorical** constraints on the relation between conjuncts

# Discussion



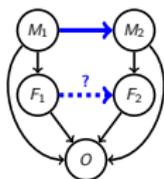
- ▶ Grammar has very little to say about **categorical** constraints on the relation between conjuncts
- ▶ But corpus data suggest there's much more to say about the **gradient** constraints on their relation

# Discussion



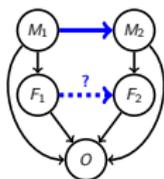
- ▶ Grammar has very little to say about **categorical** constraints on the relation between conjuncts
- ▶ But corpus data suggest there's much more to say about the **gradient** constraints on their relation
- ▶ We now have the technical tools to formally characterize these gradient constraints

# Discussion



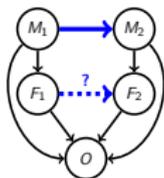
- ▶ Grammar has very little to say about **categorical** constraints on the relation between conjuncts
- ▶ But corpus data suggest there's much more to say about the **gradient** constraints on their relation
- ▶ We now have the technical tools to formally characterize these gradient constraints
- ▶ This formalization revealed a weakness of (sparse) corpus data and guided experiments to test for and quantify the strength of these constraints

# Discussion



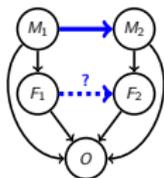
- ▶ Grammar has very little to say about **categorical** constraints on the relation between conjuncts
- ▶ But corpus data suggest there's much more to say about the **gradient** constraints on their relation
- ▶ We now have the technical tools to formally characterize these gradient constraints
- ▶ This formalization revealed a weakness of (sparse) corpus data and guided experiments to test for and quantify the strength of these constraints
- ▶ We found that gradient “Conjoin Likes” is real, and has greater explanatory reach than was ever claimed for the categorical version!

# Conjoin Likes as probabilistic grammatical knowledge



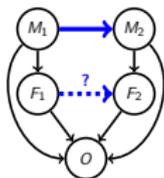
- ▶ Content of gradient *Conjoin Likes*:

# Conjoin Likes as probabilistic grammatical knowledge



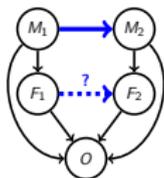
- ▶ Content of gradient *Conjoin Likes*:
  - ▶ Pointwise mutual information (pMI) is higher the more structurally similar the form of the conjuncts

# Conjoin Likes as probabilistic grammatical knowledge



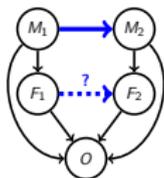
- ▶ Content of gradient *Conjoin Likes*:
  - ▶ Pointwise mutual information (pMI) is higher the more structurally similar the form of the conjuncts
  - ▶ Sensitivity of pMI to conjunct similarity falls off as structural locality decreases

# Conjoin Likes as probabilistic grammatical knowledge



- ▶ Content of gradient *Conjoin Likes*:
  - ▶ Pointwise mutual information (pMI) is higher the more structurally similar the form of the conjuncts
  - ▶ Sensitivity of pMI to conjunct similarity falls off as structural locality decreases
- ▶ These generalizations cannot be reduced to real-world knowledge or independently motivated performance constraints

# Conjoin Likes as probabilistic grammatical knowledge



- ▶ Content of gradient *Conjoin Likes*:
  - ▶ Pointwise mutual information (pMI) is higher the more structurally similar the form of the conjuncts
  - ▶ Sensitivity of pMI to conjunct similarity falls off as structural locality decreases
- ▶ These generalizations cannot be reduced to real-world knowledge or independently motivated performance constraints
- ▶ Gradient *Conjoin Likes*—and perhaps other unlexicalized and lexicalized syntactic preferences—constitutes fundamentally probabilistic grammatical knowledge

# References I

- Bayer, S. (1996). The coordination of unlike categories. *Language*, 72(3):579–616.
- Bayer, S. and Johnson, M. (1995). Features and agreement. In *Proceedings of the 1995 ACL*, pages 70–76. Association of Computational Linguistics.
- Bod, R., Hay, J., and Jannedy, S., editors (2003). *Probabilistic Linguistics*. MIT Press.
- Bresnan, J., Cueni, A., Nikitina, T., and Baayen, H. (2007). Predicting the dative alternation. In Boume, G., Kraemer, I., and Zwarts, J., editors, *Cognitive Foundations of Interpretation*, pages 69–95. Amsterdam: Royal Netherlands Academy of Science.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Clark, A. and Eyraud, R. (2006). Learning auxiliary fronting with grammatical inference. In *Proceedings of CoNLL*, pages 125–132.

## References II

- Clark, A. and Lappin, S. (2009). Another look at indirect negative evidence. In *Proceedings of the Workshop on Cognitive Aspects of Computational Language Acquisition*.
- Cooper, W. and Ross, J. (1975). World order. In *Proceedings of the Chicago Linguistic Society*, pages 63–111.
- Cuetos, F. and Mitchell, D. C. (1988). Cross-linguistic differences in parsing: Restrictions on the use of the Late Closure strategy in Spanish. *Cognition*, 30(1):73–105.
- Dalrymple, M. and Kaplan, R. (2000). Feature indeterminacy and feature resolution. *Language*, 76(4):759–798.
- Daniels, M. W. (2001). On a type-based analysis of feature neutrality and the coordination of unlikes. In *Proceedings of HPSG*, volume 8.
- Desmet, T., Brysbaert, M., and de Baecke, C. (2002). The correspondence between sentence production and corpus frequencies in modifier attachment. *The Quarterly Journal of Experimental Psychology*, 55A:879–896.

## References III

- Dubey, A., Keller, F., and Sturt, P. (2008). A probabilistic corpus-based model of syntactic parallelism. *Cognition*, 109(3):326–344.
- Frazier, L., Taft, L., Roeper, T., Clifton, C., and Ehrlich, K. (1984). Parallel structure: A source of facilitation in sentence comprehension. *Memory & Cognition*, 12(5):421–430.
- Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, 68:1–76.
- Gold, E. M. (1967). Language identification in the limit. *Information and Control*, 10:447–474.
- Hale, J., Shafran, I., Yung, L., Dorr, B., Harper, M., Krasnyanskaya, A., Lease, M., Liu, Y., Roark, B., Snover, M., et al. (2006). Pcfgs with syntactic and prosodic indicators of speech repairs. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*, pages 161–168. Association for Computational Linguistics.
- Hawkins, J. A. (1994). *A Performance Theory of Order and Constituency*. Cambridge.

## References IV

- Horning, J. (1969). *A study of grammatical inference*. PhD thesis, Stanford University.
- Ingria, R. J. P. (1990). The limits of unification. In *Proceedings of the 28th Annual Meeting of the ACL*, pages 194–204. Association for Computational Linguistics.
- Johnson, M. (1998). PCFG models of linguistic tree representations. *Computational Linguistics*, 24(4):613–632.
- Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science*, 20(2):137–194.
- Jurafsky, D. (2003). Probabilistic modeling in psycholinguistics: Linguistic comprehension and production. In Bod et al. (2003), pages 39–95.
- Klein, D. and Manning, C. D. (2003). Accurate unlexicalized parsing. In *Proceedings of ACL*.
- Levy, R. (2001). Feature indeterminacy and the coordination of unlikes in a totally well-typed HPSG. Manuscript, Stanford University.

## References V

- Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, 106(3):1126–1177.
- Levy, R. and Pollard, C. (2001). Coordination and neutralization in HPSG. In Eynde, F. V., Hellan, L., and Beermann, D., editors, *Proceedings of the 8th International Conference on Head-Driven Phrase Structure Grammar*, pages 221–234. CSLI Publications. Available online at <http://csli-publications.stanford.edu/HPSG/2/>.
- MacDonald, M. C. and Thornton, R. (2009). When language comprehension reflects production constraints: Resolving ambiguities with the help of past experience. *Memory & Cognition*, 37(8):1177–1186.
- Manning, C. D. (2003). Probabilistic syntax. In Bod et al. (2003), pages 289–341.
- Perfors, A., Tenenbaum, J. B., and Regier, T. (2011). The learnability of abstract syntactic principles. *Cognition*, 118(3):306–338.

## References VI

- Przepiórkowski, A. (1999). *Case Assignment and the Complement-Adjunct Dichotomy: A Non-Configurational Constraint-Based Approach*. PhD thesis, Universität Tübingen, Germany.
- Russell, S. and Norvig, P. (2003). *Artificial Intelligence: a Modern Approach*. Prentice Hall, second edition.
- Sag, I. (2003). Coordination and underspecification. In Kim, J.-B. and Wechsler, S., editors, *Proceedings of HPSG*, volume 10.
- Sag, I. A., Gazdar, G., Wasow, T., and Weisler, S. (1985). Coordination and how to distinguish categories. *Natural Language and Linguistic Theory*, 3:117–171.