

# The Internal Structure of Coordinate Categories

Roger Levy

UC San Diego  
Department of Linguistics  
Presented at Structure and Evidence in Linguistics

29 April 2013

# The historical arc of constraint-based grammar

- ▶ Hallmark principles of the Sag tradition of grammatical analysis:

# The historical arc of constraint-based grammar

- ▶ Hallmark principles of the Sag tradition of grammatical analysis:
  - ▶ It's essential to state formal claims about the nature of grammar precisely

# The historical arc of constraint-based grammar

- ▶ Hallmark principles of the Sag tradition of grammatical analysis:
  - ▶ It's essential to state formal claims about the nature of grammar precisely
  - ▶ We must seriously engage with the full range of distributional generalizations in the data

# The historical arc of constraint-based grammar

- ▶ Hallmark principles of the Sag tradition of grammatical analysis:
  - ▶ It's essential to state formal claims about the nature of grammar precisely
  - ▶ We must seriously engage with the full range of distributional generalizations in the data
  - ▶ But we must also be rigorous in determining when to attribute a distributional generalization to the grammar proper

# The historical arc of constraint-based grammar

- ▶ Hallmark principles of the Sag tradition of grammatical analysis:
  - ▶ It's essential to state formal claims about the nature of grammar precisely
  - ▶ We must seriously engage with the full range of distributional generalizations in the data
  - ▶ But we must also be rigorous in determining when to attribute a distributional generalization to the grammar proper
- ▶ When implemented correctly, these principles are powerful in identifying both grammatical knowledge and its interface with the rest of cognition

# 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)*

- ▶ And for case-marking (Przepiórkowski, 1999; Levy, 2001):

*proždal* “waited” governs ACC or GEN

Včera vec’ den’ on proždal [<sub>NP</sub> svoju  
 yesterday all day he expected.ACC\_OR\_GEN self’s.ACC  
 podругu Irinu] i [<sub>NP</sub> zvonka ot svoego brata  
 girlfriend.ACC Irina.ACC and call.GEN from self’s brother  
 Grigorija]. (Russian, Levy, 2001)

Gregory

“Yesterday he waited all day for his girlfriend Irina and for a call from his brother Gregory.”

# What's left for grammar?

- ▶ Generalization: a coordination is CATEGORICALLY GRAMMATICAL iff it satisfies all the *extrinsic* constraints on its well-formedness (Ingria, 1990; Bayer and Johnson, 1995; Bayer, 1996; Dalrymple and Kaplan, 2000; Daniels, 2001; Levy, 2001; Levy and Pollard, 2001; Sag, 2003)

# What's left for grammar?

- ▶ Generalization: a coordination is CATEGORICALLY GRAMMATICAL iff it satisfies all the *extrinsic* constraints on its well-formedness (Ingria, 1990; Bayer and Johnson, 1995; Bayer, 1996; Dalrymple and Kaplan, 2000; Daniels, 2001; Levy, 2001; Levy and Pollard, 2001; Sag, 2003)
- ▶ So... was “Conjoin Likes” just wrong?

# What's left for grammar?

- ▶ Generalization: a coordination is CATEGORICALLY GRAMMATICAL iff it satisfies all the *extrinsic* constraints on its well-formedness (Ingria, 1990; Bayer and Johnson, 1995; Bayer, 1996; Dalrymple and Kaplan, 2000; Daniels, 2001; Levy, 2001; Levy and Pollard, 2001; Sag, 2003)
- ▶ 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?

# 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*

(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?

- ▶ It is tempting to claim immediately that this pattern illustrates a “soft constraint” (one of Miller’s “Usage Preferences”) toward *Conjoin Likes*

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

- ▶ It is tempting to claim immediately that this pattern illustrates a “soft constraint” (one of Miller’s “Usage Preferences”) toward *Conjoin Likes*
- ▶ But should we really attribute this to the grammar proper?



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

Critical difference between nature of **evidence** for categorical versus probabilistic/gradient grammatical theories:

## Conjoin Likes as a gradient grammatical constraint?

Critical difference between nature of **evidence** for categorical versus probabilistic/gradient grammatical theories:

- ▶ Categorical: the *possibility* of a string is sufficient to demand the grammar account for it, regardless of the extralinguistic circumstances required

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

Critical difference between nature of **evidence** for categorical versus probabilistic/gradient grammatical theories:

## Conjoin Likes as a gradient grammatical constraint?

Critical difference between nature of **evidence** for categorical versus probabilistic/gradient grammatical theories:

- ▶ Probabilistic/gradient theories: the data currency is *relative prevalence*, and one must carefully disentangle the contributions of grammar and extralinguistic circumstances

## Conjoin Likes as a gradient grammatical constraint?

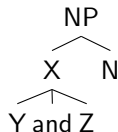
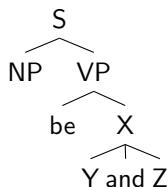
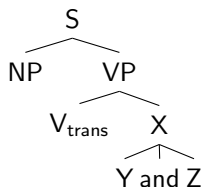
Critical difference between nature of **evidence** for categorical versus probabilistic/gradient grammatical theories:

- ▶ Probabilistic/gradient theories: the data currency is *relative prevalence*, and one must carefully disentangle the contributions of grammar and extralinguistic circumstances

## Conjoin Likes as a gradient grammatical constraint?

Critical difference between nature of **evidence** for categorical versus probabilistic/gradient grammatical theories:

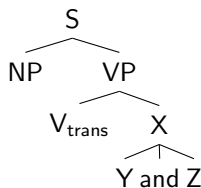
- ▶ Probabilistic/gradient theories: the data currency is *relative prevalence*, and one must carefully disentangle the contributions of grammar and extralinguistic circumstances



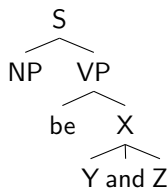
## Conjoin Likes as a gradient grammatical constraint?

Critical difference between nature of **evidence** for categorical versus probabilistic/gradient grammatical theories:

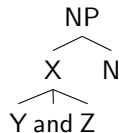
- ▶ Probabilistic/gradient theories: the data currency is *relative prevalence*, and one must carefully disentangle the contributions of grammar and extralinguistic circumstances



↓  
NP and NP



↓  
Uncorrelated Mixture

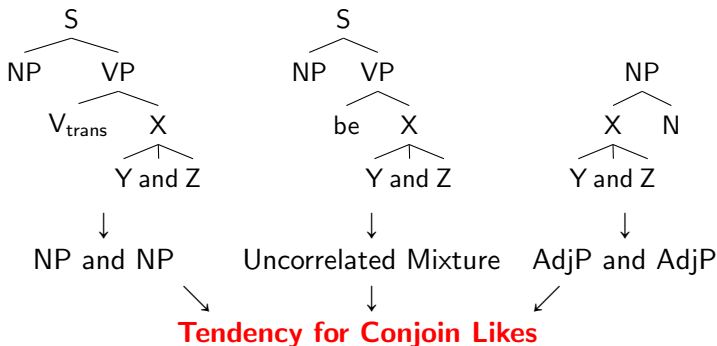


↓  
AdjP and AdjP

# Conjoin Likes as a gradient grammatical constraint?

Critical difference between nature of **evidence** for categorical versus probabilistic/gradient grammatical theories:

- ▶ Probabilistic/gradient theories: the data currency is *relative prevalence*, and one must carefully disentangle the contributions of grammar and extralinguistic circumstances

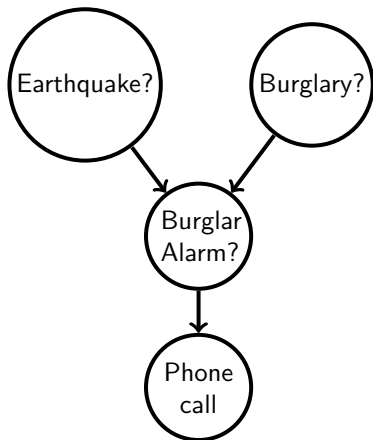




# Directed Acyclic Graphical Models (“Bayes Nets”)

Bayes Nets specify:

- ▶ Probabilistic conditional independencies:  $X$  and  $Y$  are **CONDITIONALLY INDEPENDENT** given known variables iff every path between  $X$  and  $Y$  is blocked by:

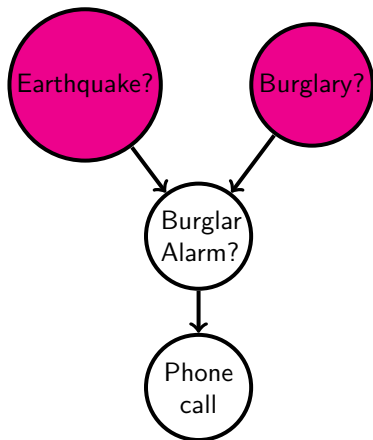


(Example due to Russell and Norvig, 2003)

# Directed Acyclic Graphical Models (“Bayes Nets”)

Bayes Nets specify:

- ▶ Probabilistic conditional independencies:  $X$  and  $Y$  are **CONDITIONALLY INDEPENDENT** given known variables iff every path between  $X$  and  $Y$  is blocked by:
  - ▶ an unknown variable **with** “converging arrows”; or

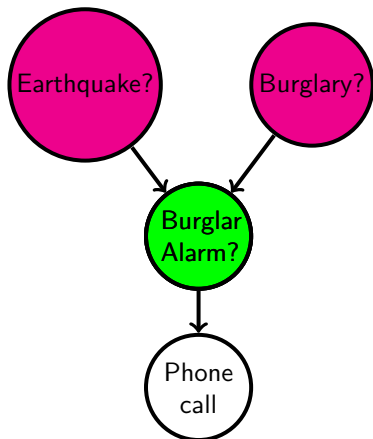


(Example due to Russell and Norvig, 2003)

# Directed Acyclic Graphical Models (“Bayes Nets”)

Bayes Nets specify:

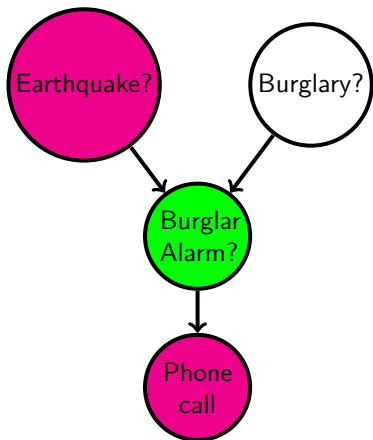
- ▶ Probabilistic conditional independencies:  $X$  and  $Y$  are **CONDITIONALLY INDEPENDENT** given known variables iff every path between  $X$  and  $Y$  is blocked by:
  - ▶ an unknown variable **with** “converging arrows”; or



(Example due to Russell and Norvig, 2003)

# Directed Acyclic Graphical Models (“Bayes Nets”)

Bayes Nets specify:

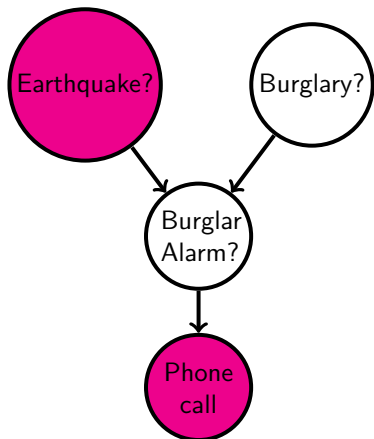


- ▶ Probabilistic conditional independencies:  $X$  and  $Y$  are **CONDITIONALLY INDEPENDENT** given known variables iff every path between  $X$  and  $Y$  is blocked by:
  - ▶ an unknown variable **with** “converging arrows”; or
  - ▶ a known variable **without** “converging arrows”

(Example due to Russell and Norvig, 2003)

# Directed Acyclic Graphical Models (“Bayes Nets”)

Bayes Nets specify:

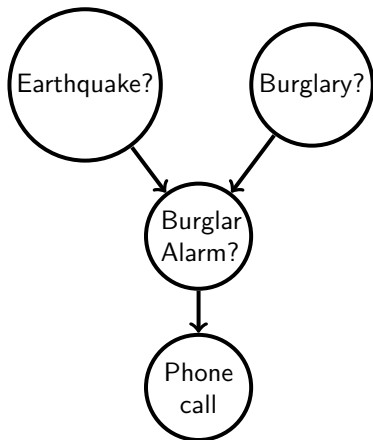


- ▶ Probabilistic conditional independencies:  $X$  and  $Y$  are **CONDITIONALLY INDEPENDENT** given known variables iff every path between  $X$  and  $Y$  is blocked by:
  - ▶ an unknown variable **with** “converging arrows”; or
  - ▶ a known variable **without** “converging arrows”

(Example due to Russell and Norvig, 2003)

# Directed Acyclic Graphical Models (“Bayes Nets”)

Bayes Nets specify:



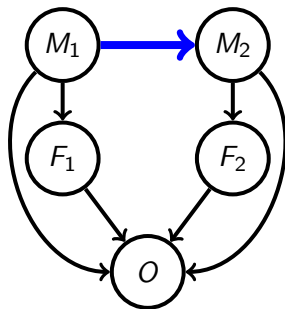
(Example due to Russell and Norvig, 2003)

- ▶ Probabilistic conditional independencies:  $X$  and  $Y$  are **CONDITIONALLY INDEPENDENT** given known variables iff every path between  $X$  and  $Y$  is blocked by:
  - ▶ an unknown variable **with** “converging arrows”; or
  - ▶ a known variable **without** “converging arrows”
- ▶ The basic units of probabilistic (=gradient) knowledge,  $P(\text{child}|\text{parents})$ :

$$P(\text{Alarm}|\text{Earthquake, Burglary})$$

$$P(\text{Call}|\text{Alarm})$$

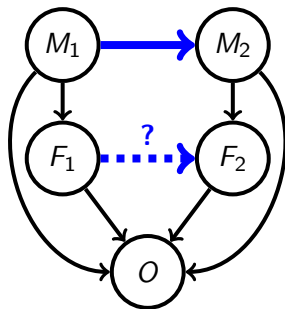
# Conjoin Likes in a probabilistic grammar



- $M_1, M_2$  Intended conjunct meanings and extrinsic constraints
- $F_1, F_2$  Realized linguistic forms of the conjuncts
- $O$  Ordering decision

(NB: Connections from  $M_i$  to  $O$  are necessary to account for semantic interpretive constraints pertaining to order, e.g., *eat and run*  $\neq$  *run and eat*; Cooper and Ross, 1975)

# Conjoin Likes in a probabilistic grammar



- $M_1, M_2$  Intended conjunct meanings and extrinsic constraints
- $F_1, F_2$  Realized linguistic forms of the conjuncts
- $O$  Ordering decision

(NB: Connections from  $M_i$  to  $O$  are necessary to account for semantic interpretive constraints pertaining to order, e.g., *eat and run*  $\neq$  *run and eat*; Cooper and Ross, 1975)



# 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

$$X \rightarrow X \text{ and } X$$

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 forms are gradiently “more grammatical” to the native speaker when they are more probable. . .

# Empirical prediction

If forms are gradiently “more grammatical” to the native speaker when they are more probable. . .

. . . then like-category coordinations should be judged to be more natural, or acceptable, than unlike-category coordinations

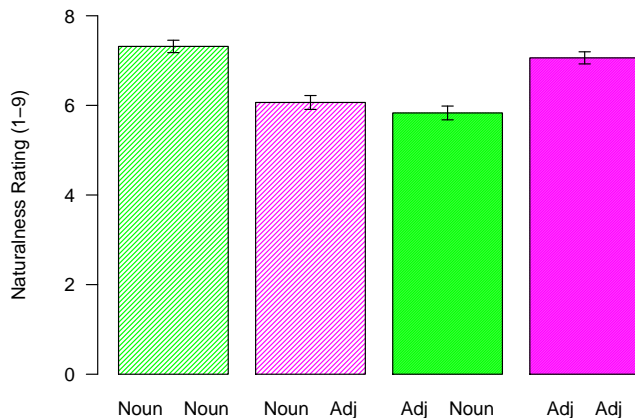
# Experiment 1

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.*)

# Experiment 1: Results



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

# Greater explanatory power of gradient constraints

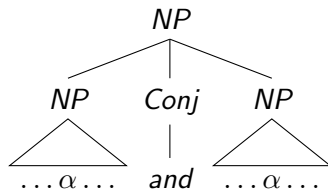
- ▶ We saw that “Conjoin Likes” is categorically false, but “probabilistically” true

# 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

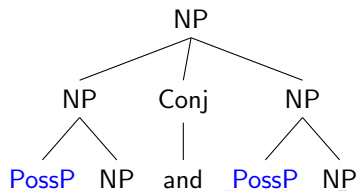
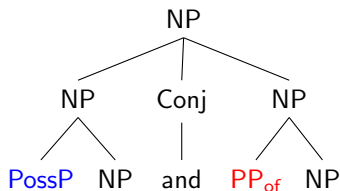
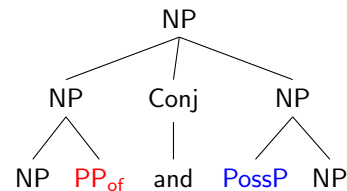
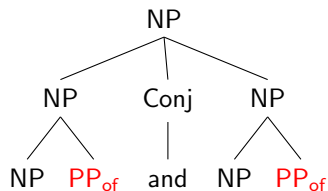
The tail of a cat

Prenominal

~ Our country's future

~ The lamp's base

~ A cat's tail



# 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 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 conjunct meanings  $M_1, M_2$
- ▶ We can control this more tightly with an experiment

# Experiment 2

Acceptability judgment study (scale of 1–9):

*Terry assembled...*

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

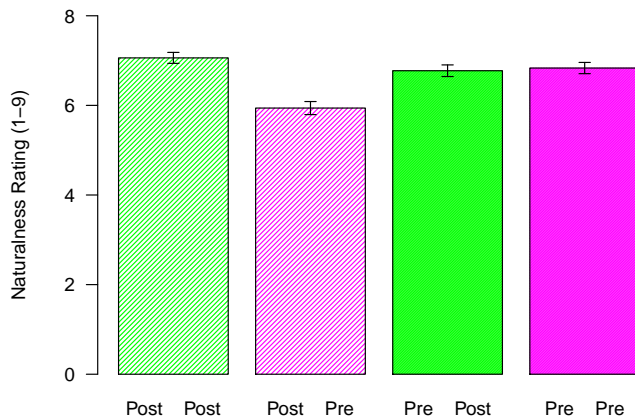
*... the frame of the chair and the lamp's base.* [Post Pre ]

*... 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.*)

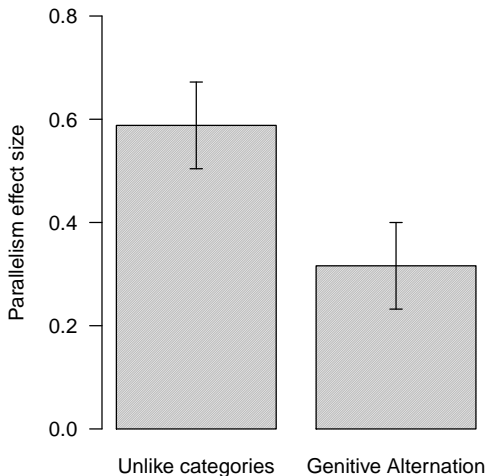
# Experiment 2: Results



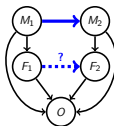
*There is also a preference for parallelism among realizations of the genitive alternation!*

# Comparison of the parallelism effects

**But Conjoin Likes > genitive parallelism!**



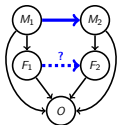
# Discussion



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

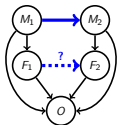


# Discussion



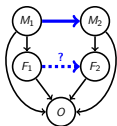
- ▶ Grammar has very little to say about **categorical** constraints on the relation between conjuncts
- ▶ But corpus data suggest there's much more to say about the **gradient** constraints on their relation

# Discussion



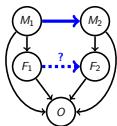
- ▶ Grammar has very little to say about **categorical** constraints on the relation between conjuncts
- ▶ But corpus data suggest there's much more to say about the **gradient** constraints on their relation
- ▶ We now have the technical tools to formally characterize these gradient constraints

# 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!

# Discussion

But why should gradient “Conjoin Likes” exist in the first place?

# Discussion

But why should gradient “Conjoin Likes” exist in the first place?

- ▶ Stylistic preference?

# Discussion

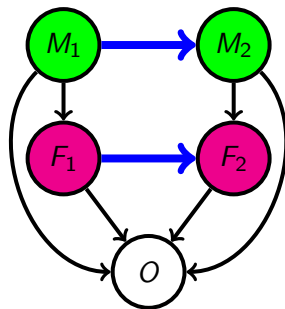
But why should gradient “Conjoin Likes” exist in the first place?

- ▶ Stylistic preference?
- ▶ Psychological mechanism (priming)?

# Discussion

But why should gradient “Conjoin Likes” exist in the first place?

- ▶ Stylistic preference?
- ▶ Psychological mechanism (priming)?
- ▶ Or is there a deeper source of explanation in the nature of probabilistic grammatical knowledge?

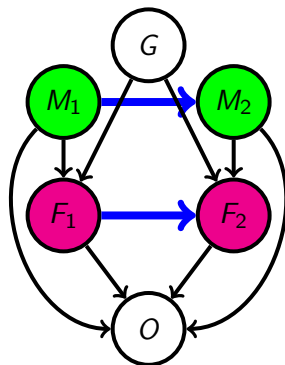




# Discussion

But why should gradient “Conjoin Likes” exist in the first place?

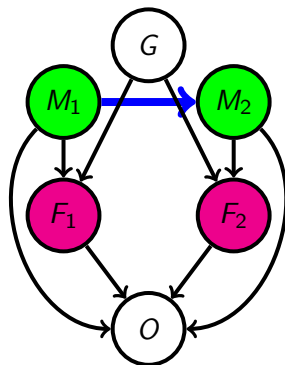
- ▶ Stylistic preference?
- ▶ Psychological mechanism (priming)?
- ▶ Or is there a deeper source of explanation in the nature of probabilistic grammatical knowledge?



# Discussion

But why should gradient “Conjoin Likes” exist in the first place?

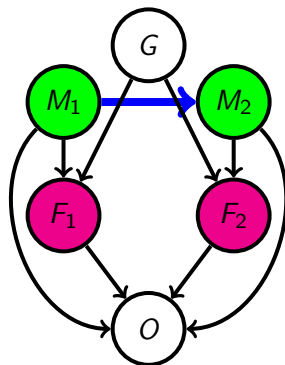
- ▶ Stylistic preference?
- ▶ Psychological mechanism (priming)?
- ▶ Or is there a deeper source of explanation in the nature of probabilistic grammatical knowledge?



# Discussion

But why should gradient “Conjoin Likes” exist in the first place?

- ▶ Stylistic preference?
- ▶ Psychological mechanism (priming)?
- ▶ Or is there a deeper source of explanation in the nature of probabilistic grammatical knowledge?
- ▶ The Sag tradition of **precise formal claims**, **serious engagement with data**, and **rigor in assigning credit for distributional generalizations** will be essential to working this out



# Thank you, Ivan!

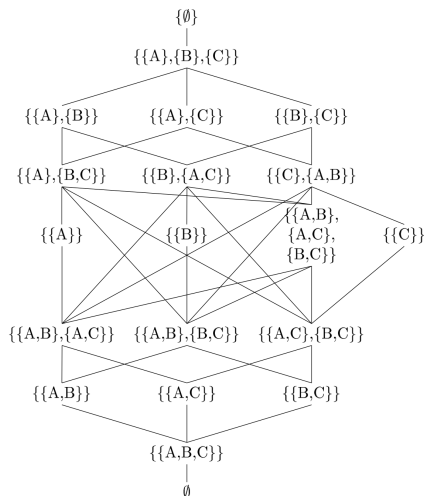


Figure 5: Double-set lattice ordered by  $\supseteq$  over  $\{A, B, C\}$ .

# 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.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.
- Cooper, W. and Ross, J. (1975). World order. In *Proceedings of the Chicago Linguistic Society*, pages 63–111.
- 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.
- Dubey, A., Keller, F., and Sturt, P. (2008). A probabilistic corpus-based model of syntactic parallelism. *Cognition*, 109(3):326–344.

## References II

- 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.
- 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.
- 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.
- Klein, D. and Manning, C. D. (2003). Accurate unlexicalized parsing. In *Proceedings of ACL*.

## References III

- Levy, R. (2001). Feature indeterminacy and the coordination of unlikes in a totally well-typed HPSG. MS., Stanford University.
- Levy, R. and Pollard, C. (2001). Coordination and neutralization in HPSG. In *Proceedings of the 8th International Conference on Head-Driven Phrase Structure Grammar*. CSLI Publications.
- 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.