# Simulating Structural Priming Effects with PIPS



Zhenghao "Herbert" Zhou, Robert Frank Yale University

SCiL Presentation Jun 15, 2023

# Simulating Structural Priming Effects with PIPS



Zhenghao "Herbert" Zhou, Robert Frank Yale University

SCiL Presentation Jun 15, 2023

#### Parallelism in Producing Syntax [Brehm et al. 2022]

#### Structural Priming

- Structural priming effect;
- Lexical boost effect (LBE);
- Inverse frequency effect (IFE);

#### Structural Priming

#### Computational Modeling

- Structural priming effect;
- Lexical boost effect (LBE);
- Inverse frequency effect (IFE);

The PIPS Model

The Gradient Symbolic Computation (GSC) framework

Structural Priming	Computational Modeling	Current Study
• Structural priming effect;	<ul><li>The PIPS Model</li><li>The Gradient</li></ul>	• How to model the priming

- Lexical boost effect (LBE);
- Inverse frequency effect (IFE);

- Symbolic Computation (GSC) framework
- procedure?
- How to quantify the three priming effects?

Structural Priming	Computational Modeling	Current Study	Results & Discussion
<ul> <li>Structural priming effect;</li> <li>Lexical boost effect (LBE);</li> <li>Inverse frequency effect (IFE);</li> </ul>	<ul> <li>The PIPS Model</li> <li>The Gradient Symbolic Computation (GSC) framework</li> </ul>	<ul> <li>How to model the priming procedure?</li> <li>How to quantify the three priming effects?</li> </ul>	• PIPS is capable of qualitatively capturing the structural priming effects.

## Structural Priming

**Structural Priming:** speakers tend to reuse the syntactic structures they have recently encountered during production.



## Structural Priming

**Structural Priming:** speakers tend to reuse the syntactic structures they have recently encountered during production.

**Our focus**: Double Object (DO) vs. Prepositional Dative (PD) for ditransitive predicates.

## Structural Priming

**Structural Priming:** speakers tend to reuse the syntactic structures they have recently encountered during production.

**Our focus**: Double Object (DO) vs. Prepositional Dative (PD) for ditransitive predicates.

- DO: Alice sent Bob a letter.
- PD: Alice sent a letter to Bob.

E.g. [Bock 1986]



E.g. [Bock 1986, Bernolet & Hartsuiker 2010]

Alice sent Bob a letter. [DO] -<u>.00</u>

#### PRIME

Production Task with <u>Preamble</u> <u>Completion</u> <u>Paradigm</u>

E.g. [Bock 1986, Bernolet & Hartsuiker 2010]



E.g. [Bock 1986, Bernolet & Hartsuiker 2010]



Does lexical selection play any role?

Does lexical selection play any role?

**Lexical Boost Effect:** structural priming effect is stronger when the word that heads the primed structures is repeated between prime and target sentences.

Does lexical selection play any role?

**Lexical Boost Effect:** structural priming effect is stronger when the word that heads the primed structures is repeated between prime and target sentences.

Alice <u>gave</u> Bob a book.

TARGET

Does lexical selection play any role?

**Lexical Boost Effect:** structural priming effect is stronger when the word that heads the primed structures is repeated between prime and target sentences.

Carl <u>gave</u> Danis a letter.

PRIME

Carl showed Danis a letter.

Alice <u>gave</u> Bob a book.

TARGET

Do verb biases play any role?



Verb Bias: show is biased towards PD

Do verb biases play any role?



**Inverse Frequency Effect:** the less preferred syntactic structures cause stronger priming effect than the more preferred structures.

Do verb biases play any role?



**Inverse Frequency Effect:** the less preferred syntactic structures cause stronger priming effect than the more preferred structures.

Alice showed Bob a letter. [DO]

Alice showed a letter to Bob. [PD]

PRIME

Do verb biases play any role?



**Inverse Frequency Effect:** the less preferred syntactic structures cause stronger priming effect than the more preferred structures.

Alice showed Bob a letter. [DO]

Alice showed a letter to Bob. [PD]

PRIME

Carl gave Danis a book. [DO]

Carl gave a book to Danis. [PD] TARGET

# Interim Summary

- Structural Priming Effect (StrucPriming)
- Lexical Boost Effect (LBE)
- Inverse Frequency Effect (IFE)

# **Interim Summary**

- Structural Priming Effect (StrucPriming)
- Lexical Boost Effect (LBE)
- Inverse Frequency Effect (IFE)

# Next: computational modeling

- Gradient Symbolic Computation (GSC)
- Parallelism in Producing Syntax (PIPS)

#### Gradient Symbolic Computation Framework

A GSC Parser: a continuous-state, continuous-time stochastic dynamical-system model of symbolic processing.

• Motivation: integrating the *discrete* (symbolic) and *continuous* (gradient) aspects of language processing that happen at different levels of the mind.

[Cho & Smolensky 2018, 2020, Smolensky & Hale 2006]

### Gradient Symbolic Computation Framework

**A GSC Parser:** a continuous-state, continuous-time stochastic dynamical-system model of symbolic processing.

• Motivation: integrating the *discrete* (symbolic) and *continuous* (gradient) aspects of language processing that happen at different levels of the mind.

#### **Core Properties**:

- Representing discrete symbolic structures in a continuous space.
- Maintaining multiple locally-coherent structures simultaneously at each timestep; gradually converging to a final, discrete structure / hypothesis.
- Implementing Harmonic Grammar to measure well-formedness of structures.

[Cho & Smolensky 2018, 2020, Smolensky & Hale 2006]

#### **Tensor Product Representation**

# **Orthogonal vector representations of the followings:**

- **Roles**: positional, structural information;
- **Fillers**: the values that are filled into the roles;
- **Binding**: outer/tensor product of roles and fillers;

[Smolensky 1990]

#### **Tensor Product Representation**

# Orthogonal vector representations of the followings:

- **Roles**: positional, structural information;
- **Fillers**: the values that are filled into the roles;
- **Binding**: outer/tensor product of roles and fillers;



[Smolensky 1990]

### **Tensor Product Representation**

# **Orthogonal vector representations of the followings:**

- **Roles**: positional, structural information;
- **Fillers**: the values that are filled into the roles;
- **Binding**: outer/tensor product of roles and fillers;



 $f_{
m S} \otimes r_{
m root} + f_{
m NP} \otimes r_0 + f_{
m VP} \otimes r_1 + f_{
m V} \otimes r_{10} + f_{
m PP} \otimes r_{11}$ 

[Smolensky 1990]

**Motivation:** modeling *sentence production* with the GSC properties, with a case study in modeling *agreement attraction*.

**Motivation:** modeling *sentence production* with the GSC properties, with a case study in modeling *agreement attraction*.

**PREAMBLE**: *The key to the cabinets...* 

[Brehm et al. 2022, Bock & Miller 1991]

**Motivation:** modeling *sentence production* with the GSC properties, with a case study in modeling *agreement attraction*.

## PREAMBLE: *The key* to *the cabinets*... Subject Intervener



[Brehm et al. 2022, Bock & Miller 1991]

**Motivation:** modeling *sentence production* with the GSC properties, with a case study in modeling *agreement attraction*.





Method: representing sentence production as a Preamble Completion Task

- Representing preambles in memory as pre-activations of bindings;
- Dynamically and stochastically modeling the transient activations of a blend of alternative structures (which lead to errors in production) also for structural priming. [Brehm et al. 2022, Bock & Miller 1991]

# Interim Summary

- Gradient Symbolic Computation
- Parallelism in Producing Syntax

# Interim Summary

- Gradient Symbolic Computation
- Parallelism in Producing Syntax

# Next: current study

- How to model priming?
- How to quantify Structural Priming, LBE, and IFE?

#### Training: Fillers and Verb Biases

[Pickering & Branigan 1998, Yi et al. 2019]
### Training: Fillers and Verb Biases

Nine ditransitive verbs with their absolute frequencies of DO vs. PD:

Verb	DO Frequency	PD Frequency
give	15311	8402
show	502	571
send	658	3134
lend	177	677
hand	308	659
loan	12	11
offer	752	1203
sell	190	1288
post	1	55

[Pickering & Branigan 1998, Yi et al. 2019]

### Training: Fillers and Verb Biases

Nine ditransitive verbs with their absolute frequencies of DO vs. PD:

**Abstracting away Noun Phrases:** 

• NP<sub>s</sub>, NP<sub>i</sub>, NP<sub>d</sub> are the three noun phrases;

Verb	DO Frequency	PD Frequency
give	15311	8402
show	502	571
send	658	3134
lend	177	677
hand	308	659
loan	12	11
offer	752	1203
sell	190	1288
post	1	55

[Pickering & Branigan 1998, Yi et al. 2019]

### Training: Fillers and Verb Biases

Nine ditransitive verbs with their absolute frequencies of DO vs. PD:

**Abstracting away Noun Phrases:** 

• NP<sub>s</sub>, NP<sub>i</sub>, NP<sub>d</sub> are the three noun phrases;

Two types of grammatical sentences:

- **DO:** NP<sub>s</sub> \_VERB\_ NP<sub>i</sub> NP<sub>d</sub>
- **PD:** NP<sub>s</sub> \_VERB\_ NP<sub>d</sub> to NP<sub>i</sub>

Verb	DO Frequency	PD Frequency
give	15311	8402
show	502	571
send	658	3134
lend	177	677
hand	308	659
loan	12	11
offer	752	1203
sell	190	1288
post	1	55

[Pickering & Branigan 1998, Yi et al. 2019]

#### ABS







ABS





**NORM** 

#### Structure

DO PD

ABS





**NORM** 

#### BASE



ABS





NORM

#### BASE



Training: minimizing KL-divergence between production distribution and PCFG distribution.

Preamble Activation: raise the activation value of the bindings to 0.5;
Priming: raise the activation value of the bindings to {0.05, 0.1, 0.2};

Preamble Activation: raise the activation value of the bindings to 0.5;Priming: raise the activation value of the bindings to {0.05, 0.1, 0.2};

**Priming Modes:** 



Structure

Preamble Activation: raise the activation value of the bindings to 0.5;Priming: raise the activation value of the bindings to {0.05, 0.1, 0.2};

**Priming Modes:** 



Structure

Words

NPi

Preamble Activation: raise the activation value of the bindings to 0.5;Priming: raise the activation value of the bindings to {0.05, 0.1, 0.2};

**Priming Modes:** 



For each model:

1. <u>Prime</u> the model with a verb+structure combination. [18 conditions in total]



#### For each model:

- 1. <u>Prime</u> the model with a verb+structure combination. [18 conditions in total]
- 2. Activate the preamble: NP<sub>S</sub> VERB.



#### For each model:

- 1. <u>Prime</u> the model with a verb+structure combination. [18 conditions in total]
- 2. Activate the preamble: NP<sub>s</sub> VERB.
- 3. Run the model for 50 production trials as <u>target</u> productions.



#### For each model:

- 1. <u>Prime</u> the model with a verb+structure combination. [18 conditions in total]
- 2. Activate the preamble: NP<sub>s</sub> VERB.
- 3. Run the model for 50 production trials as <u>target</u> productions.
- 4. Obtain the proportion of production of the18 grammatical sentences + *Others*.



Sample production trials for preamble NP<sub>s</sub> offer.

#### For each model:

- 1. <u>Prime</u> the model with a verb+structure combination. [18 conditions in total]
- 2. Activate the preamble: NP<sub>s</sub> VERB.
- 3. Run the model for 50 production trials as <u>target</u> productions.
- 4. Obtain the proportion of production of the18 grammatical sentences + *Others*.

**Baseline Production for Comparison**: run the model without priming and obtain the proportions of the 18 sentences.



## Quantifying Priming Effects

**Target Verb:** the verb given in the preamble that should be produced. We focus on the portion of produced sentences with correct target verbs.

## Quantifying Priming Effects

**Target Verb:** the verb given in the preamble that should be produced. We focus on the portion of produced sentences with correct target verbs.

**DO Ratio:** 
$$\operatorname{Ratio}_v(\operatorname{DO}) = \frac{\#\operatorname{DO}_v}{\#\operatorname{DO}_v + \#\operatorname{PD}_v}$$

## Quantifying Priming Effects

**Target Verb:** the verb given in the preamble that should be produced. We focus on the portion of produced sentences with correct target verbs.

**DO Ratio:** 
$$\operatorname{Ratio}_{v}(\operatorname{DO}) = \frac{\#\operatorname{DO}_{v}}{\#\operatorname{DO}_{v} + \#\operatorname{PD}_{v}}$$

**Deviation from Baseline:** 

 $\operatorname{Dev}_v^{v'}(\operatorname{DO}) = \operatorname{Ratio}_v(\operatorname{DO})_{\operatorname{primed by} v'} - \operatorname{Ratio}_v(\operatorname{DO})_{\operatorname{unprimed}}$ 

### Quantifying Priming Effects, cont.

# $ext{StrucPriming} \propto \sum_{v \in \mathcal{V}} \sum_{s \in \{ ext{DO, PD}\}} ext{Dev}_v( ext{s})$

### Quantifying Priming Effects, cont.

$$ext{StrucPriming} \propto \sum_{v \in \mathcal{V}} \sum_{s \in \{ ext{DO, PD}\}} ext{Dev}_v( ext{s})$$

# $ext{LBE} \propto \sum_{v,v' \in \mathcal{V}} \sum_{s \in \{ ext{DO, PD}\}} [ ext{Dev}_v^v( ext{s}) - ext{Dev}_v^{v'}( ext{s})]$

### Quantifying Priming Effects, cont.

$$ext{StrucPriming} \propto \sum_{v \in \mathcal{V}} \sum_{s \in \{ ext{DO, PD}\}} ext{Dev}_v( ext{s})$$

$$ext{LBE} \propto \sum_{v,v' \in \mathcal{V}} \sum_{s \in \{ ext{DO, PD}\}} [ ext{Dev}_v^v( ext{s}) - ext{Dev}_v^{v'}( ext{s})]$$

$$ext{IFE} \propto \sum_{v \in \mathcal{V}} \sum_{v' \in \mathcal{V}_{ ext{PD}}} [ ext{Dev}_v^{v'}( ext{DO}) - ext{Dev}_v^{v'}( ext{PD})]$$

# Interim Summary

- Trained 3 models: ABS, NORM, BASE
- Priming means activating corresponding bindings
   Priming weights: {0.05, 0.1, 0.2}
  - Priming modes: {structure, words, whole}
- Quantifying effects w.r.t. deviations from baseline.

# Interim Summary

- Trained 3 models: ABS, NORM, BASE
- Priming means activating corresponding bindings
   Priming weights: {0.05, 0.1, 0.2}
  - Priming modes: {structure, words, whole}
- Quantifying effects w.r.t. deviations from baseline.

# Next: results and discussion

### Results: priming weights and modes



### Results: priming weights and modes



• Greater priming weights lead to stronger priming effect.

### Results: priming weights and modes



- Greater priming weights lead to stronger priming effect.
- Priming Modes: structure < words</li>
   <(=) whole;</li>

0.39

Whole

NORM

ABS



66

IFE

0.3

0.02

0.39

Whole

NORM

ABS



\* Priming weight = 0.2

0.06

0.05

0.1





\* Priming weight = 0.2

All three models captured LBE;





- \* Priming weight = 0.2
  - All three models captured LBE;
  - Prime by structures does not produce LBE;

0.06

0.05

0.1





- \* Priming weight = 0.2
  - All three models captured LBE;
  - Prime by structures does not produce LBE;
  - For IFE, comparing ABS and NORM models (BASE doesn't encode verb bias):

0.1





\* Priming weight = 0.2

- All three models captured LBE;
- Prime by structures does not produce LBE;
- For IFE, comparing ABS and NORM models (BASE doesn't encode verb bias);
- Only the **NORM** model captures IFE;

### Conclusions & Implications
### Conclusions & Implications

- The PIPS model is capable of qualitatively modeling human results on structural primings;
  - In general, the GSC framework works on modeling the incremental processes of sentence production.
  - It remains a question of how to quantitatively align simulation results with human results.

## **Conclusions & Implications**

- The PIPS model is capable of qualitatively modeling human results on structural primings;
  - In general, the GSC framework works on modeling the incremental processes of sentence production.
  - It remains a question of how to quantitatively align simulation results with human results.
- Why only the **NORM** model captures IFE remains a problem for further exploration.
  - The **NORM** model learns the PCFG distribution better than the **ABS** model;
  - Both models acquired most of the verb biases;
  - Postulation: PIPS may be sensitive to target distributions and/or hyperparameters.
  - Connection with the role of Type (NORM) vs. Token (ABS) Frequency (e.g. [Bybee & Hopper 2001]);

### Future Work

**Improving Current Simulations** 

- Controlling the total amount of priming activation among the 3 priming modes;
- Computing the IFE score by weighting the verb biases from individual verbs;

### Future Work

#### **Improving Current Simulations**

- Controlling the total amount of priming activation among the 3 priming modes;
- Computing the IFE score by weighting the verb biases from individual verbs;

#### **Deeper into the Theories of Structural Priming**

• Understanding the relationship between the current PIPS model and theories of structural priming: for instances, *transient activation account* (e.g. [Pickering & Branigan 1998]), *implicit learning account* (e.g. [Chang et al. 2002, 2006]);

### Future Work

#### **Improving Current Simulations**

- Controlling the total amount of priming activation among the 3 priming modes;
- Computing the IFE score by weighting the verb biases from individual verbs;

#### **Deeper into the Theories of Structural Priming**

• Understanding the relationship between the current PIPS model and theories of structural priming: for instances, *transient activation account* (e.g. [Pickering & Branigan 1998]), *implicit learning account* (e.g. [Chang et al. 2002, 2006]);

#### **Extending the Simulation Domain**

• Simulating other structural priming phenomena, such as filler-gap dependency (e.g. [Momma 2022]);

### Reference

- Sarah Bernolet and Robert J. Hartsuiker. 2010. Does verb bias modulate syntactic priming? *Cognition*, 114(3):455–461.
- J. Kathryn Bock. 1986. Syntactic persistence in language production. *Cognitive Psychology*, 18(3):355–387.
- Kathryn Bock and Carol A Miller. 1991. Broken agreement. *Cognitive Psychology*, 23(1):45–93.
- Laurel Brehm, Pyeong Whan Cho, Paul Smolensky, and Matthew A. Goldrick. 2022. PIPS: A Parallel Planning Model of Sentence Production. *Cognitive Science*, 46(2):e13079.
- Pyeong Whan Cho, Matthew Goldrick, Richard L. Lewis, and Paul Smolensky. 2018. Dynamic encoding of structural uncertainty in gradient symbols. In *Proceedings of the 8th Workshop on Cognitive Modeling and Computational Linguistics (CMCL 2018)*, pages 19–28, Salt Lake City, Utah. Association for Computational Linguistics.
- Pyeong Whan Cho, Matthew Goldrick, and Paul Smolensky. 2020. Parallel parsing in a Gradient Symbolic Computation parser.
- John Hale and Paul Smolensky. 2006. Harmonic gram- mars and harmonic parsers for formal languages. *Smolensky and Legendre*, pages 393–416.
- T. Florian Jaeger and Neal Snider. 2007. Implicit Learning and Syntactic Persistence: Surprisal and Cumulativity. *University of Rochester Working Papers in the Language Sciences*, 3:26–44.
- Martin J. Pickering and Holly P. Branigan. 1998. The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language*, 39(4):633–651.
- Eunkyung Yi, Jean-Pierre Koenig, and Douglas Roland. 2019. Semantic similarity to high-frequency verbs affects syntactic frame selection. *Cognitive Linguistics*, 30(3):601–628.

# Thanks for Listening!

## **Q&A Session**

### Tensor Product Representation: Extra

#### **Orthogonal vector representations of the followings:**

- Representation of a syntactic tree = summing over all bindings representing each node;
- **Decomposability**: linearly independence avoids the superposition catastrophe:

[Smolensky 1990]

### Tensor Product Representation: Extra

#### **Orthogonal vector representations of the followings:**

- Representation of a syntactic tree = summing over all bindings representing each node;
- **Decomposability**: linearly independence avoids the superposition catastrophe:

$$f_1+f_2=f_2+f_1$$

 $f_1\otimes r_1+f_2\otimes r_2
eq f_2\otimes r_1+f_1\otimes r_2$ 

[Smolensky 1990]

### **Brick-Role Representation**

#### **TPR Implementation: Brick-Role Representation**



[Cho 2020, Brehm et al. 2022]

### Optimizing Over Constraints

[Smolensky & Hale 2006, Cho et al. 2018, 2020]

### **Optimizing Over Constraints**

**Non-grammatical Constraints:** 

- **Competition Constraint**: avoiding bindings of multiple fillers to the same role;
- **Baseline Constraint**: avoiding extreme states;
- **Discreteness Constraint**: encouraging activation values to be close to 0 or 1 as time goes;
  - $\circ$  Commitment policy *q*

[Smolensky & Hale 2006, Cho et al. 2018, 2020]

### **Optimizing Over Constraints**

**Non-grammatical Constraints:** 

- **Competition Constraint**: avoiding bindings of multiple fillers to the same role;
- **Baseline Constraint**: avoiding extreme states;
- **Discreteness Constraint**: encouraging activation values to be close to 0 or 1 as time goes;
  - $\circ$  Commitment policy q

#### **Grammatical Constraints**:

• Harmonic Grammar



[Smolensky & Hale 2006, Cho et al. 2018, 2020]

#### PCFG for the NORM Model

 $1 S \rightarrow NPs VP$  $1 \text{ XP} \rightarrow \text{NPi NPd}$ 1 YP -> NPd PP  $1 PP \rightarrow P NPi$ 0.05848089468779124 VP -> show XP 0.06651910531220875 VP -> show YP 0.021690400843881855 VP -> send XP 0.10330959915611815 VP -> send YP 0.025907494145199064 VP -> lend XP 0.09909250585480094 VP -> lend YP 0.03981385729058945 VP -> hand XP 0.08518614270941055 VP -> hand YP 0.06521739130434782 VP -> loan XP 0.059782608695652176 VP -> loan YP 0.04808184143222506 VP -> offer XP 0.07691815856777494 VP -> offer YP 0.016069012178619755 VP -> sell XP 0.10893098782138025 VP -> sell YP 0.002232142857142857 VP -> post XP 0.12276785714285714 VP -> post YP