Syntactic Priming

Slides by Grusha Prasad

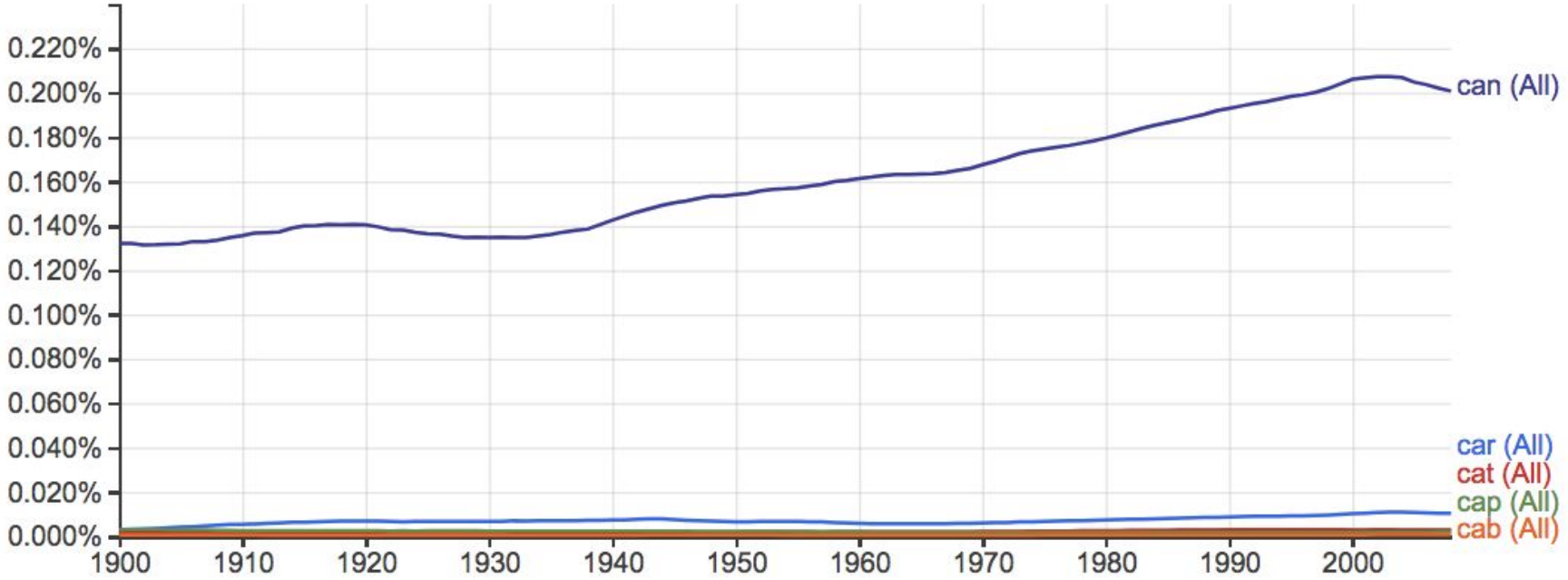
DOG

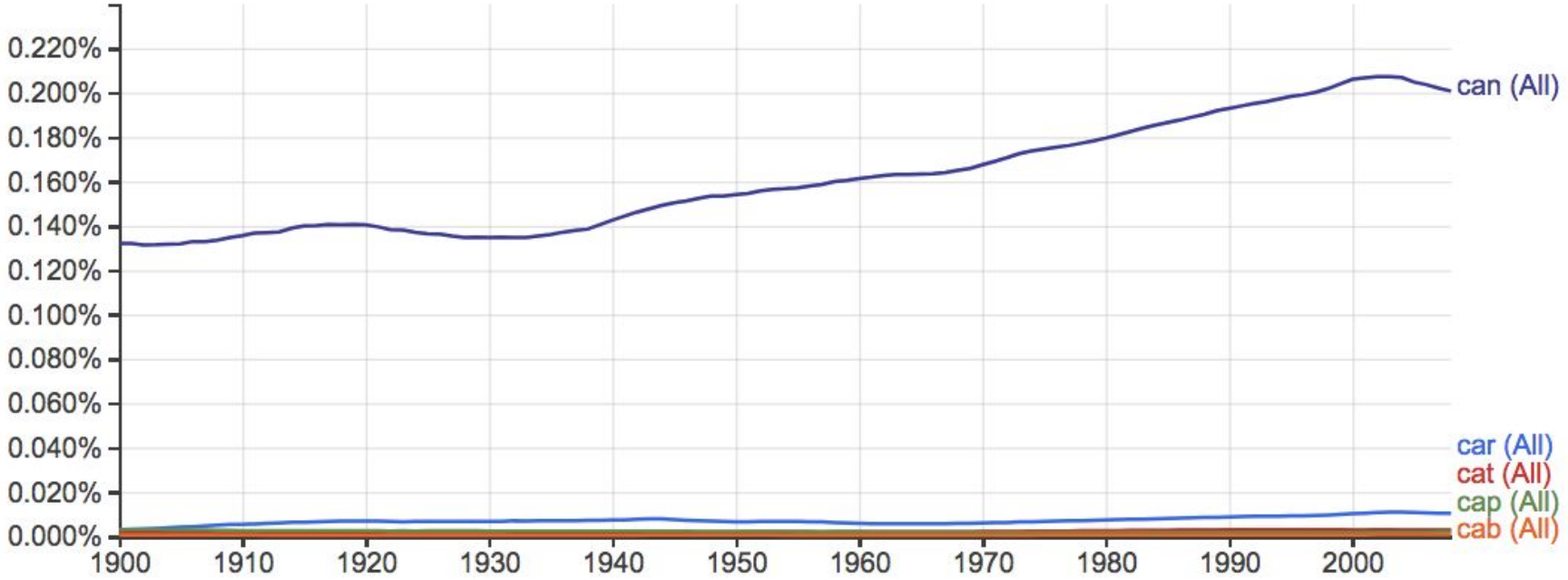
CA_

CAT

CAT

CAN CAP CAR CAB


0.220%
0.200%
0.180%
0.160%
0.140%
0.120%
0.100%
0.080%
0.060%
0.040%
0.020%
0.000%
1900
1910
1920
1930
1940
1950
1960
1970
1980
1990
2000
can (All)
car (All)
cat (All)
cap (All)
cab (All)



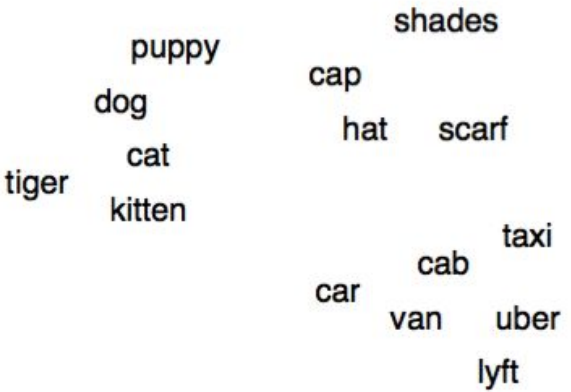
Why did we pick *cat* even if it wasn't the most probable option?

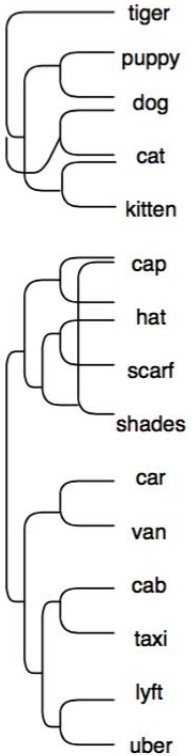
Priming

A stimulus is easier to produce/ process if it is preceded by a **related** stimulus.

Priming

A stimulus is easier to produce/ process if it is preceded by a **related** stimulus.

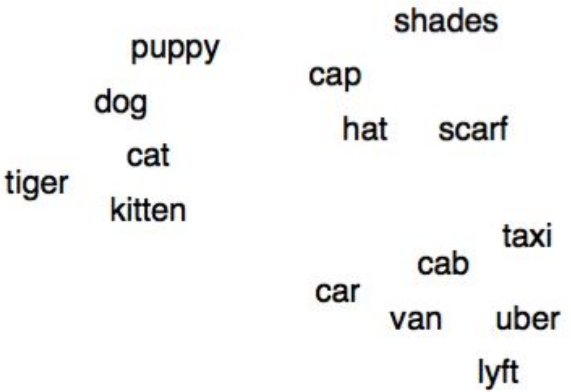


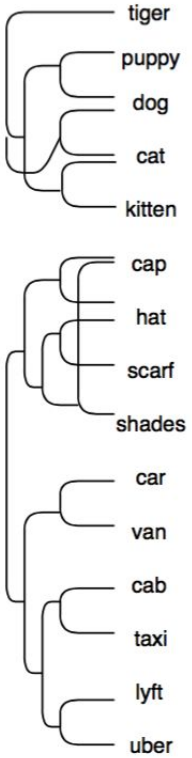


In whatever distance metric we use, *dog* and *cat* should be closer together than *dog* and *car*.

Priming

A stimulus is easier to produce/ process if it is preceded by a **related** stimulus.





In whatever distance metric we use, *dog* and *cat* should be closer together than *dog* and *car*.

Priming allows us to estimate the distance between words, sentences, images etc. Gives insight into internal representations

Outline

1. What is syntactic priming?
    a. Production
    b. Comprehension
2. Why does syntactic priming occur?
    a. Spreading activation.
        - Computational implementation: Dubey et al (2006)
    b. Implicit learning
        - Computational implementation (ish): van Schijndel & Linzen (2018)
3. Graded syntactic priming and how it can provide insights into how neural language models (e.g. RNNs) and humans represent syntax.

Outline

1. What is syntactic priming?
   a. Production
   b. Comprehension
2. Why does syntactic priming occur?
   a. Spreading activation.
      - Computational implementation: Dubey et al (2006)
   b. Implicit learning
      - Computational implementation (ish): van Schijndel & Linzen (2018)
3. Graded syntactic priming and how it can provide insights into how neural language models (e.g. RNNs) and humans represent syntax.

The woman sent the man a letter

Syntactic priming in production

Bock (1986)

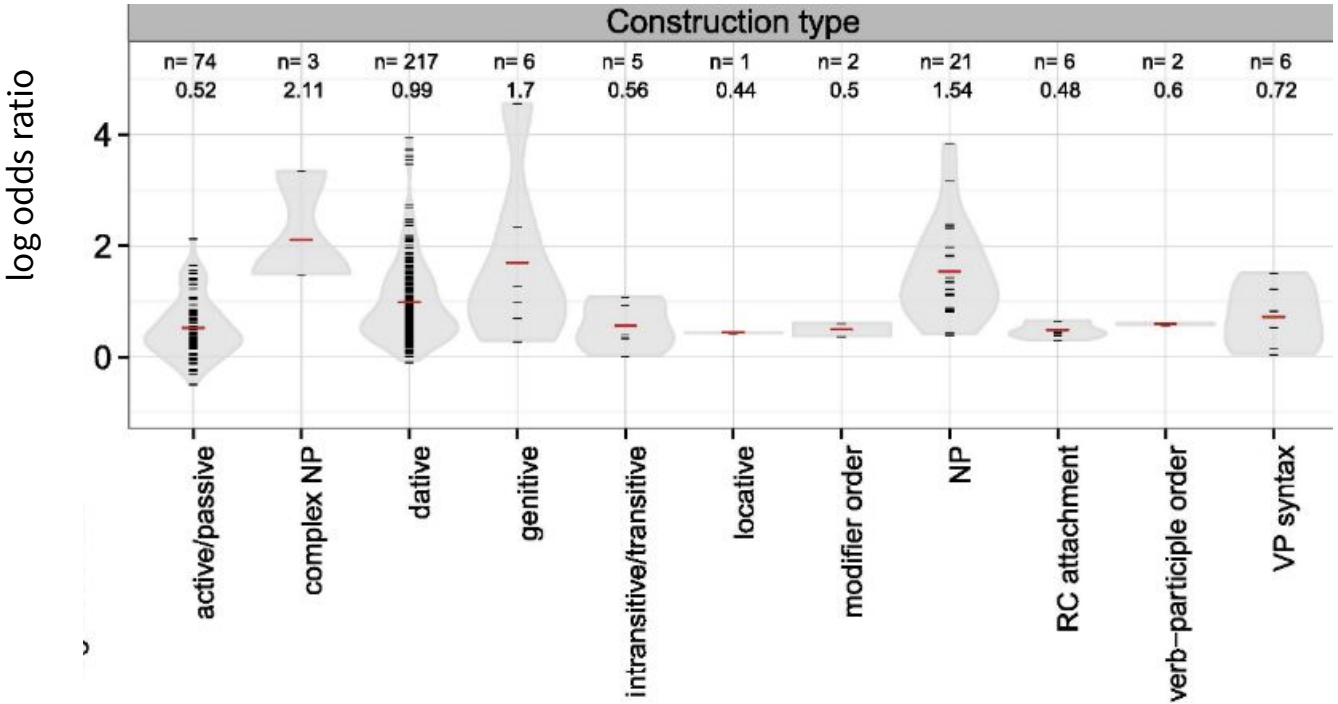
Effect of Syntactic Priming on Form of Sentences: Percentages of Utterances in Four Syntactic Forms following Priming Sentences in the Same or an Alternative Form

		Utterance form		
Priming condition		Prepositional dative	Double-object dative	Total
Prepositional dative	PO	48	31	79
Double-object dative	DO	25	53	78
Difference		23 ± (8), (10)	22 ± (7), (13)	

Syntactic priming in production

Mahowald et al (2016)

$$\text{LogOddsRatio} = \log\left(\frac{p(X|\text{Prime})}{1-p(X|\text{Prime})}\right) - \log\left(\frac{p(X|\text{NoPrime})}{1-p(X|\text{NoPrime})}\right)$$



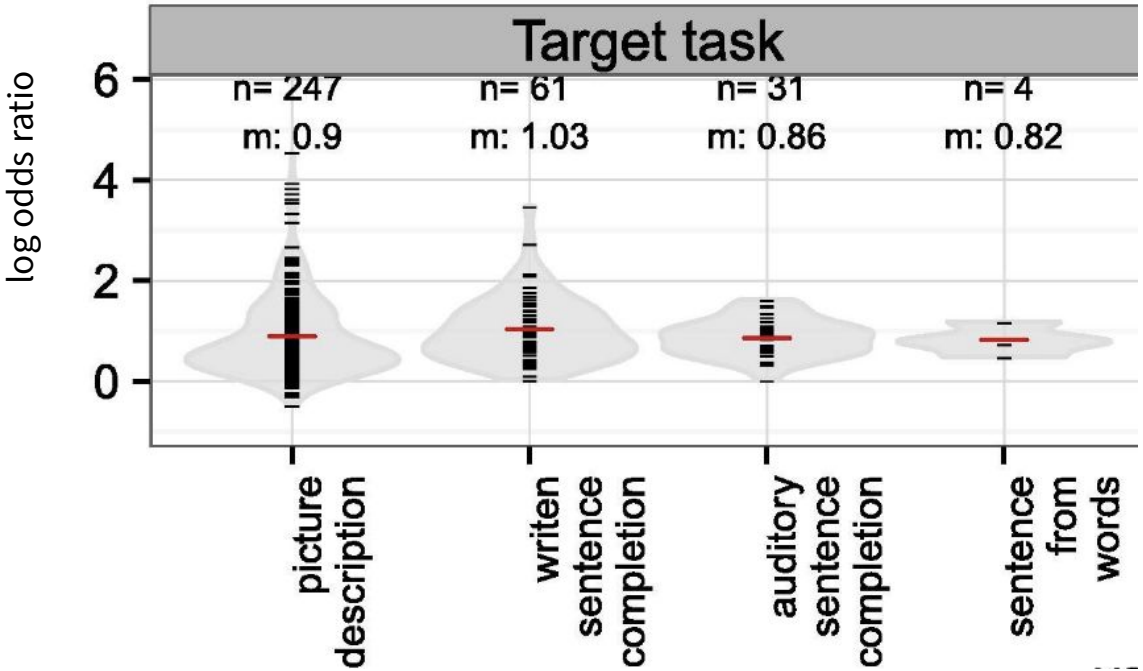
n: number of studies

Find priming for a wide range of structures

Syntactic priming in production

Mahowald et al (2016)

$$\text{LogOddsRatio} = \log\left(\frac{p(X|\text{Prime})}{1-p(X|\text{Prime})}\right) - \log\left(\frac{p(X|\text{NoPrime})}{1-p(X|\text{NoPrime})}\right)$$

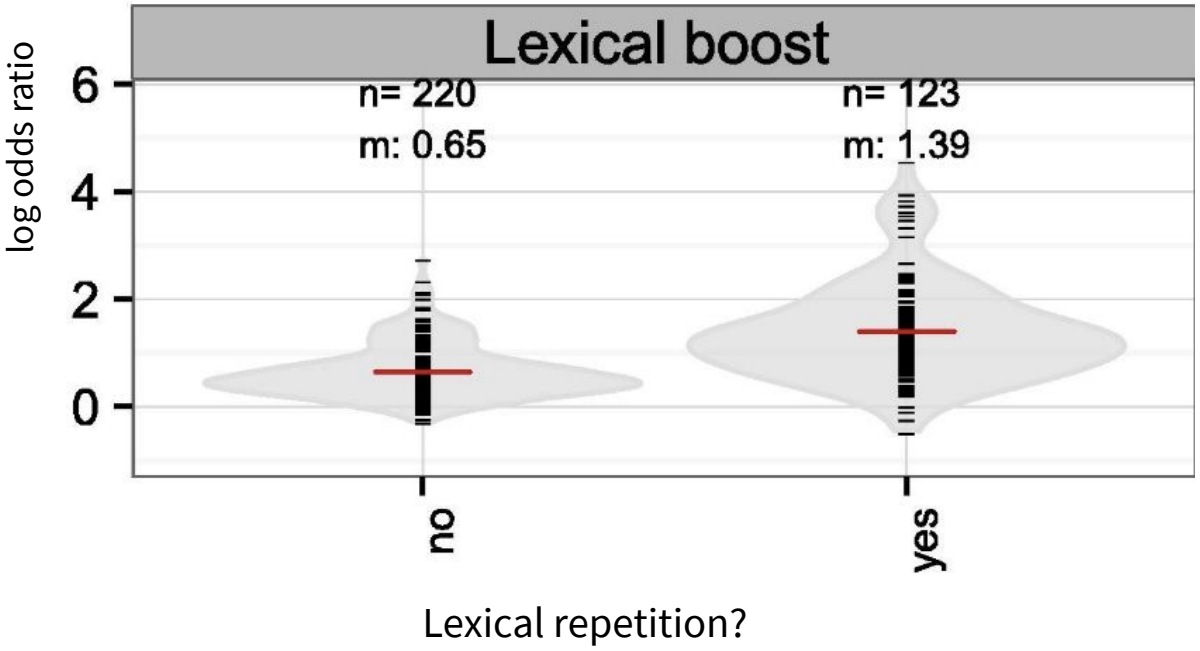


n: number of studies

Syntactic priming in production

Mahowald et al (2016)

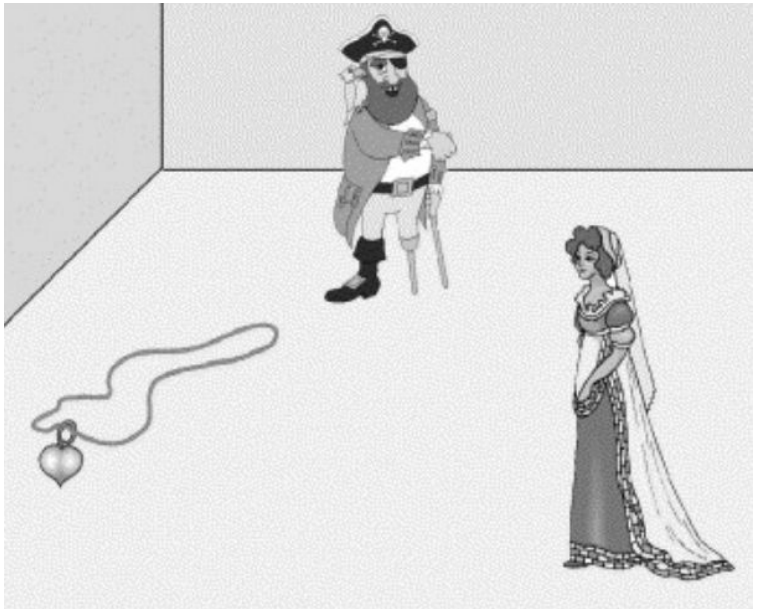
$$\text{LogOddsRatio} = \log\left(\frac{p(X|\text{Prime})}{1-p(X|\text{Prime})}\right) - \log\left(\frac{p(X|\text{NoPrime})}{1-p(X|\text{NoPrime})}\right)$$



n: number of studies

Syntactic priming in comprehension

Arai et al (2007)



DO Prime:
The assassin will send the dictator the parcel

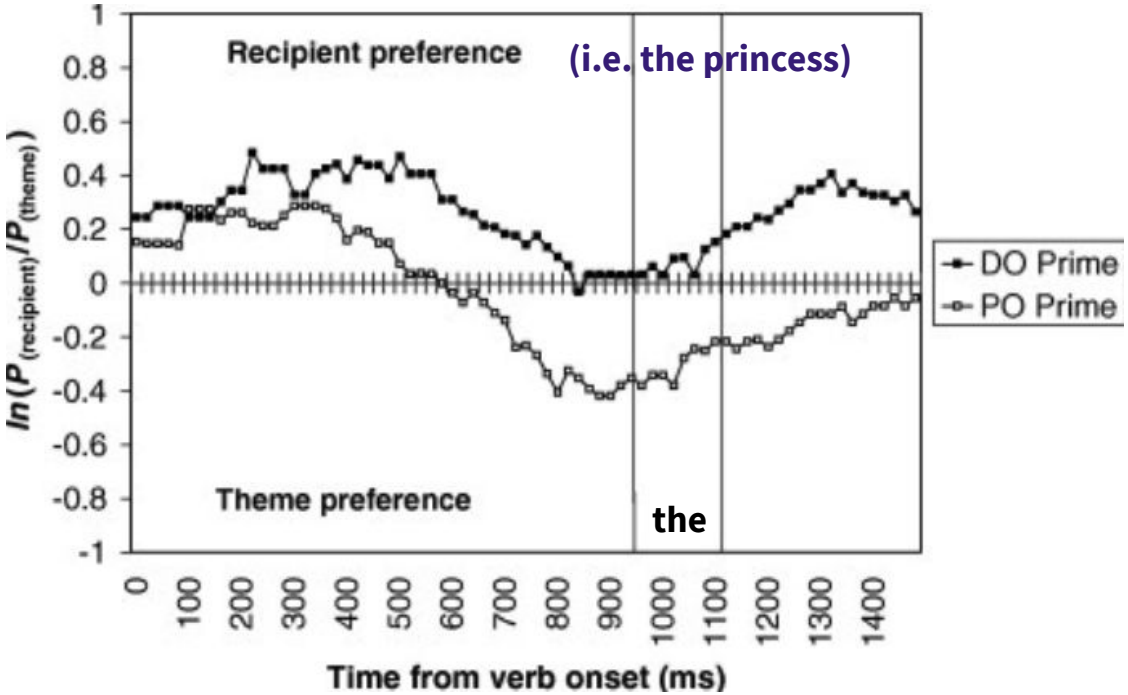
DO Target:
The pirate will send **the** princess the necklace

Similar PO prime and targets

Where will participants look when they hear *the* ?

Syntactic priming in comprehension

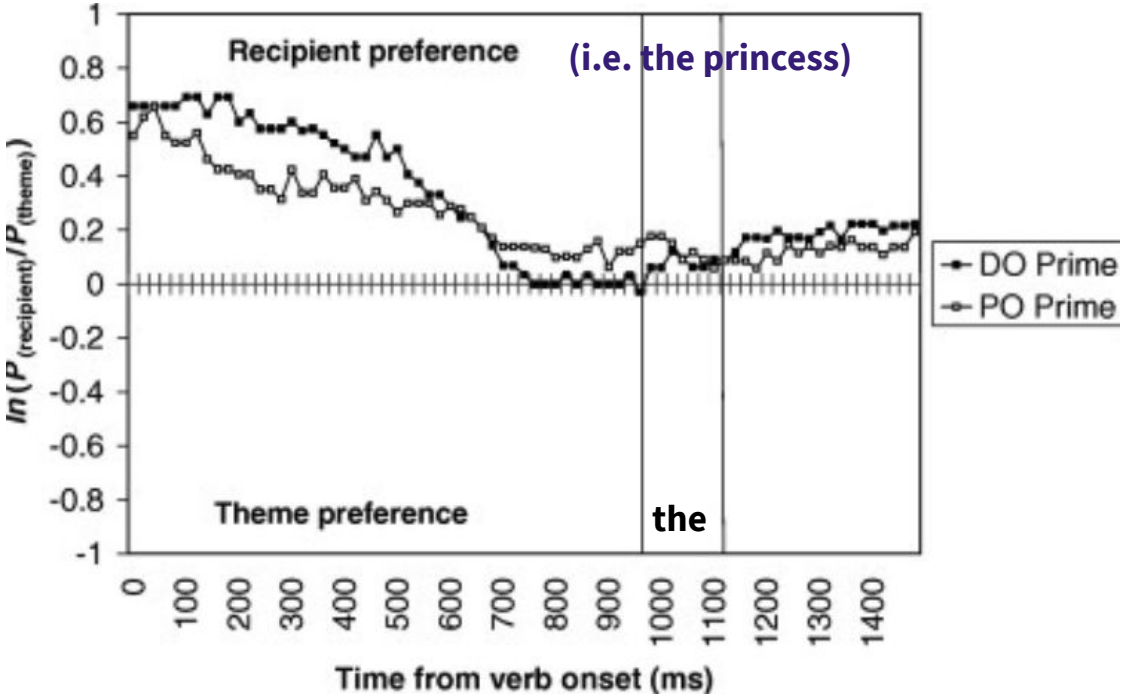
Arai et al (2007)



With verb overlap

Syntactic priming in comprehension

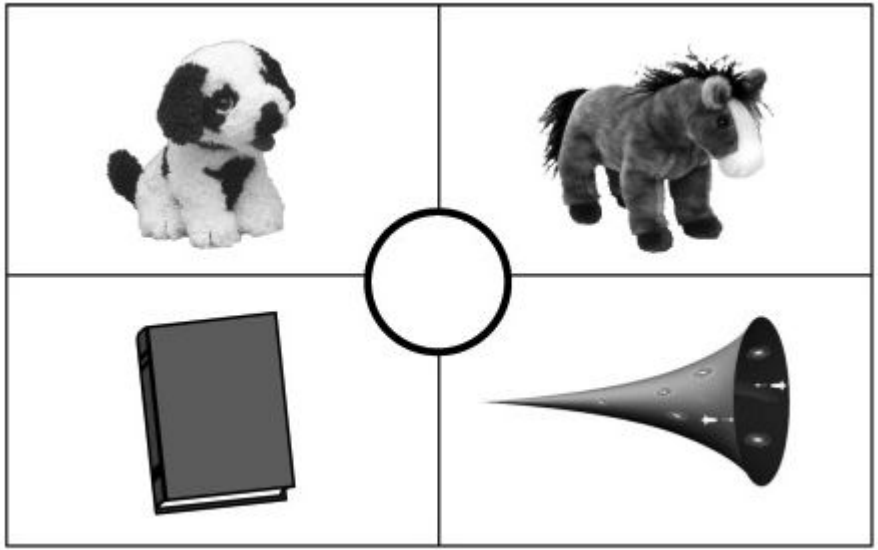
Arai et al (2007)



Without verb overlap

Syntactic priming in comprehension

Thotathiri & Snedeker (2008)



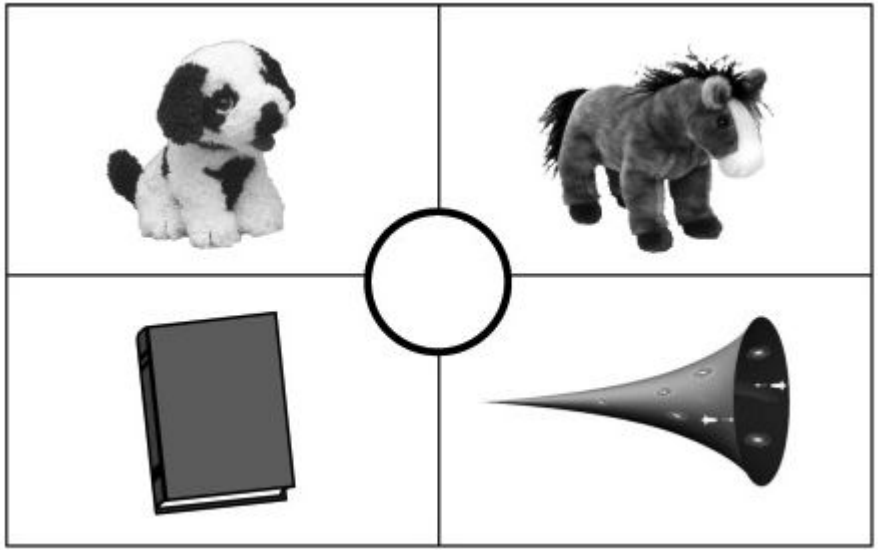
PO target: Show the horn to the horse

DO target: Show the horse the horn

What time window are we interested in?

Syntactic priming in comprehension

Thotathiri & Snedeker (2008)



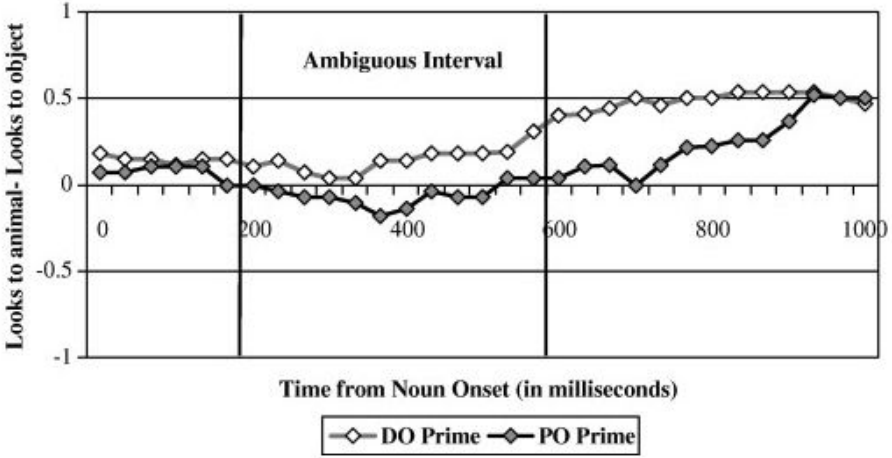
PO target: Show the **hor**n to the horse

DO target: Show the **hor**se the horn

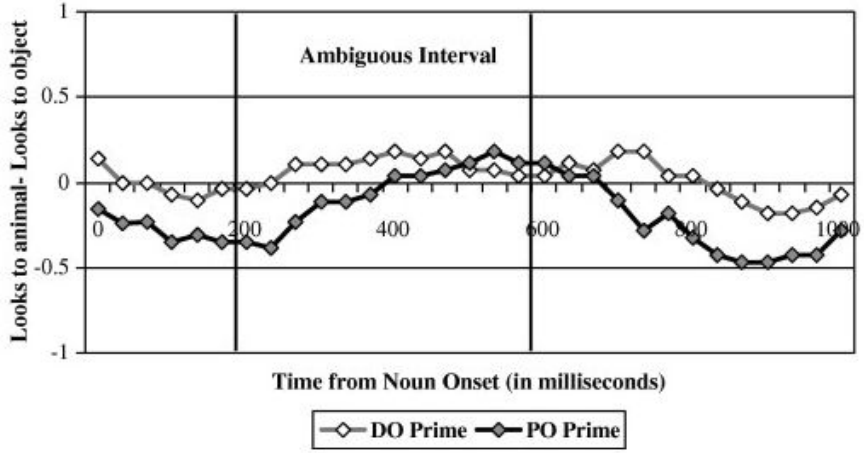
What time window are we interested in?

Syntactic priming in comprehension

Thotathiri & Snedeker (2008)



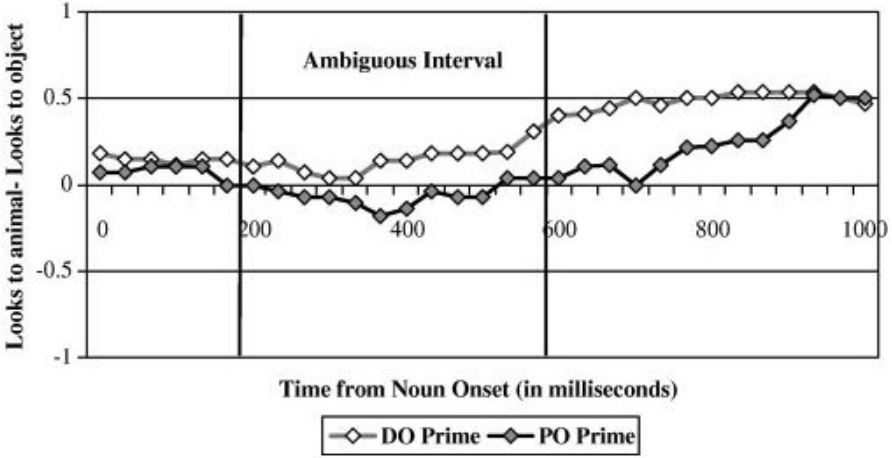
DO target

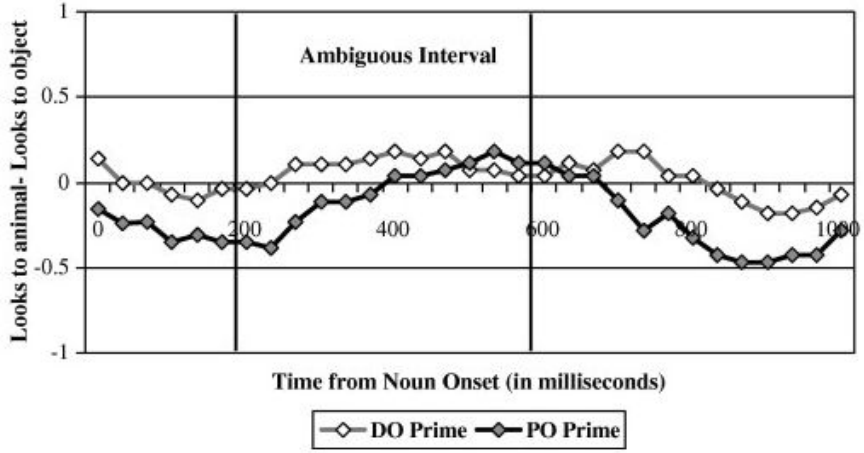


PO target

Syntactic priming in comprehension

Thotathiri & Snedeker (2008)





DO target

Note, target type does not matter in this paradigm. Why?

PO target

Syntactic priming in comprehension

Tooley et al (2006)

Repetition prime:
The man watched by the woman was tall and handsome

Synonym Prime:
The man observed by the woman was tall and handsome

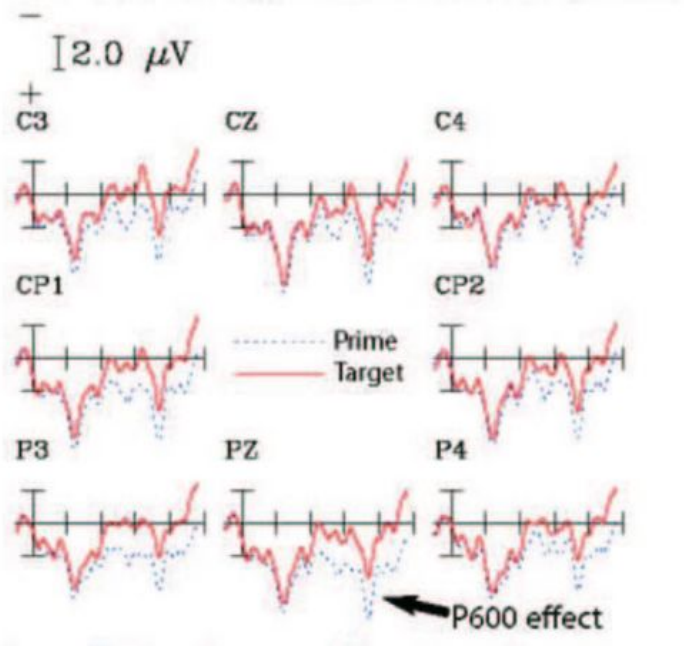
Target:
The child watched **by** the parent was playing quietly

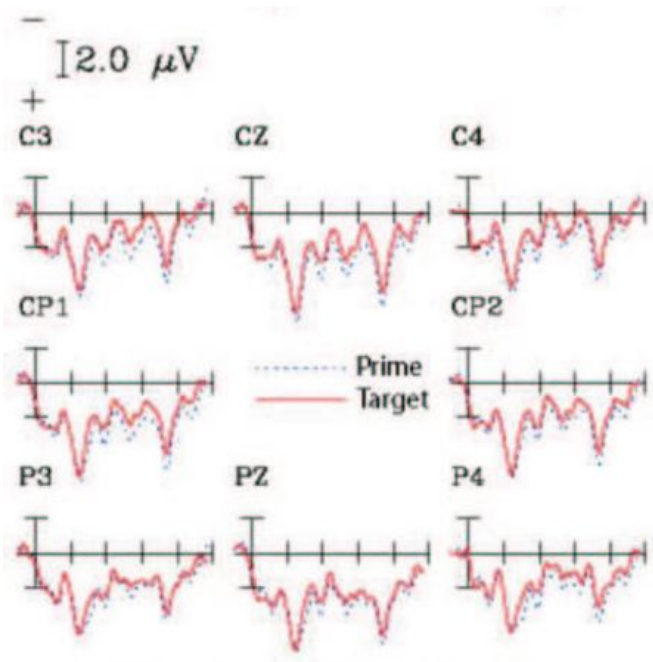
Reduced RCs lead to garden path effects.
Priming = Decrease in garden path effect

Syntactic priming in comprehension

Target: watched
Repeated: watched
Synonym: observed

Tooley et al (2009)





Repeated: Prime = Target

Synonym: Decrease in GP effect

Measuring garden path effect with P600 amplitude

Syntactic priming in comprehension

Target: watched
Repeated: watched
Synonym: observed

Tooley et al (2009)

Table 1
Mean Values of the Four Dependent Measures by Scoring Region and Condition for Experiment 2

	Regression			
Scoring region	First pass	First-pass regressions	Path time	Total time
Verb region				
Repeated baseline	309	17.5%	380	468
Repeated target	289	18.7%	353	421
Synonym baseline	312	22.7%	364	476
Synonym target	316	16.2%	371	400
PP region				
Repeated baseline	542	21.6	699	814
Repeated target	501	17.7	638	721
Synonym baseline	546	21.3	681	755
Synonym target	530	18.0	655	738

Significant

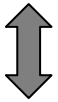
Not significant

Measuring garden path effect with reading times

But what does syntactic priming tell us?

But what does syntactic priming tell us?

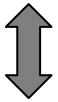
A stimulus is easier to produce/ process if it is preceded by a **related** stimulus.



If stimulus A is easier to produce/process after stimulus B, then A and B are related

But what does syntactic priming tell us?

A stimulus is easier to produce/ process if it is preceded by a **related** stimulus.



If stimulus A is easier to produce/process after stimulus B, then A and B are related

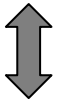
The girl handed the man a paintbrush

The rock star sold the astronomer a telescope

What is similar?

But what does syntactic priming tell us?

A stimulus is easier to produce/ process if it is preceded by a **related** stimulus.



If stimulus A is easier to produce/process after stimulus B, then A and B are related

The girl handed the man a paintbrush

The rock star sold the astronomer a telescope

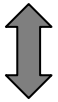
What is similar?

VP —> V NP NP
VP —> V the N a N

How can we eliminate this option?

But what does syntactic priming tell us?

A stimulus is easier to produce/ process if it is preceded by a **related** stimulus.



If stimulus A is easier to produce/process after stimulus B, then A and B are related

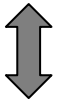
The girl handed the man a paintbrush

The rock star sold an undercover agent some cocaine

What is similar?

But what does syntactic priming tell us?

A stimulus is easier to produce/ process if it is preceded by a **related** stimulus.



If stimulus A is easier to produce/process after stimulus B, then A and B are related

The girl handed the man a paintbrush

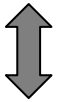
The rock star sold an undercover agent some cocaine

What is similar?

VP —> V NP NP

But what does syntactic priming tell us?

A stimulus is easier to produce/ process if it is preceded by a **related** stimulus.



If stimulus A is easier to produce/process after stimulus B, then A and B are related

The girl handed the man a paintbrush

The rock star sold an undercover agent some cocaine

What is similar?

VP —> V NP NP

VP —> V_{handed} NP NP

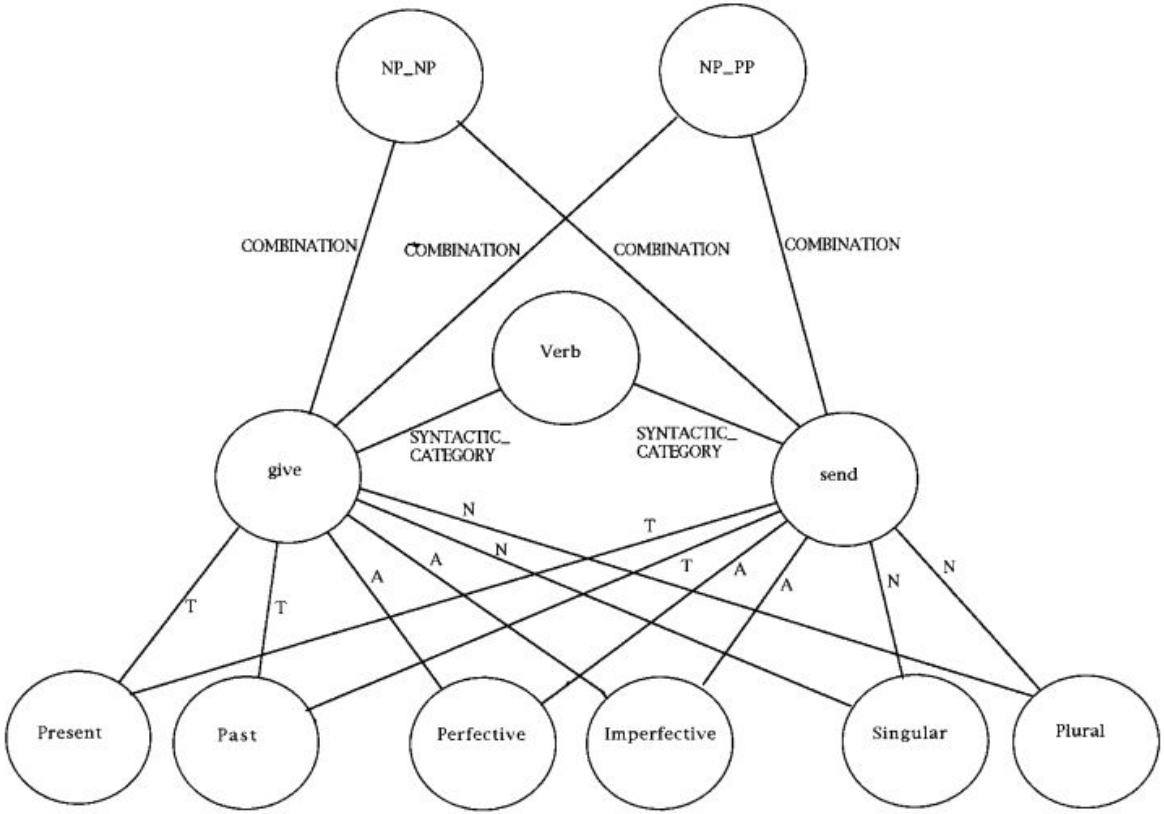
Lexical boost effect

Outline

1. What is syntactic priming?
    a. Production
    b. Comprehension
2. Why does syntactic priming occur?
    a. Spreading activation.
        - Computational implementation: Dubey et al (2006)
    b. Implicit learning
        - Computational implementation (ish): van Schijndel & Linzen (2018)
3. Graded syntactic priming and how it can provide insights into how neural language models (e.g. RNNs) and humans represent syntax.

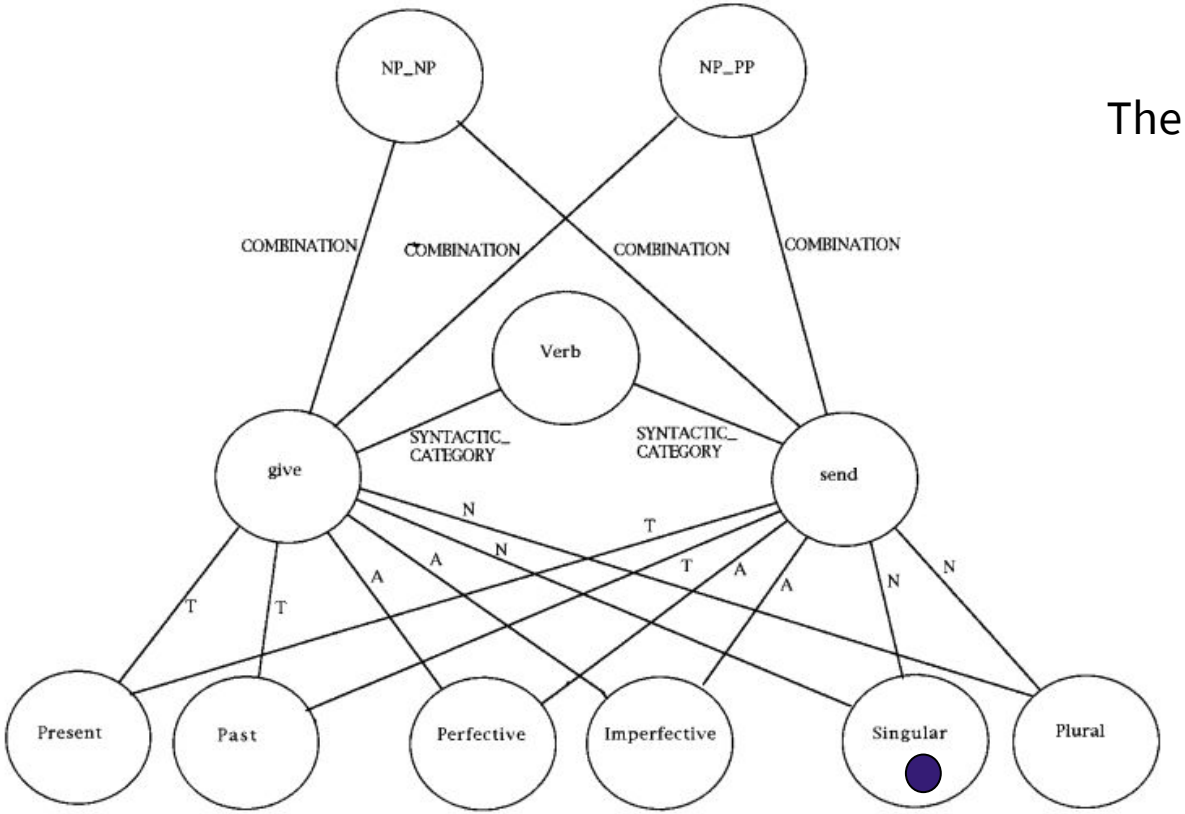
Spreading activation

Pickering & Branigan (1998)



Spreading activation

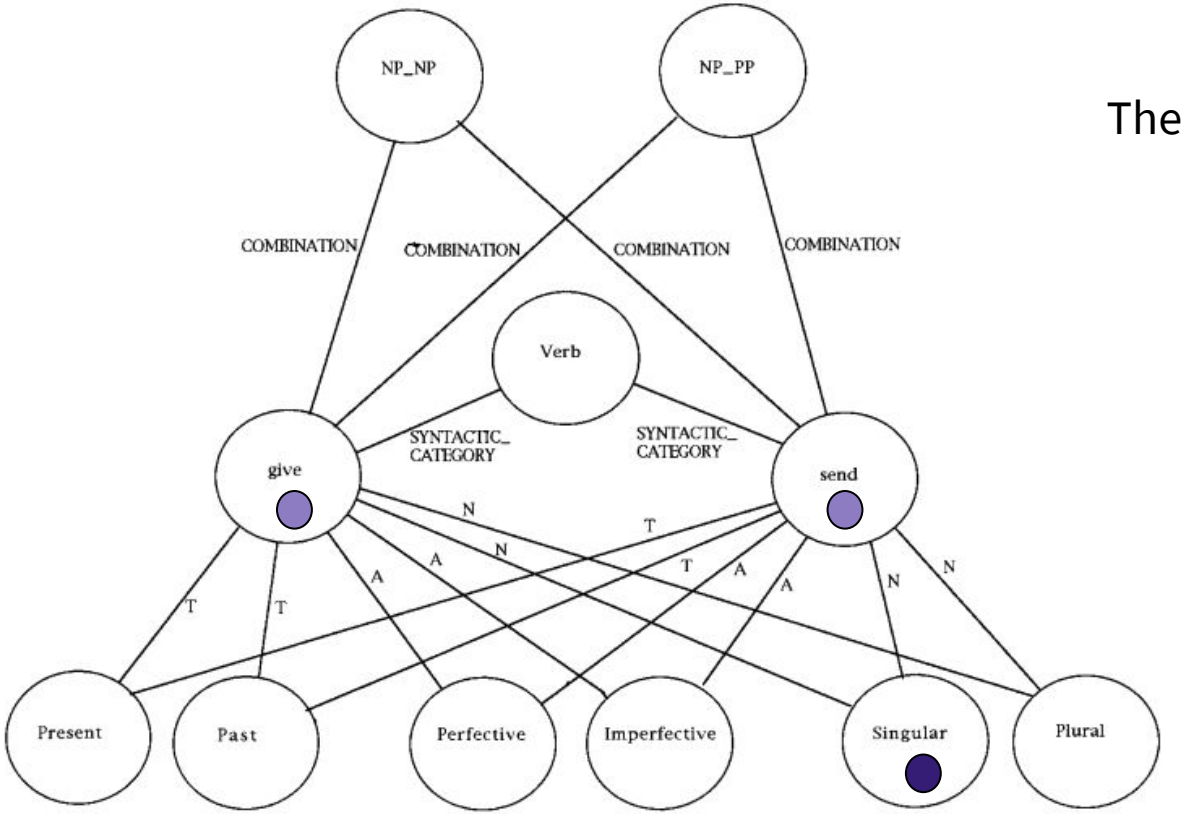
Pickering & Branigan (1998)



The girl …

Spreading activation

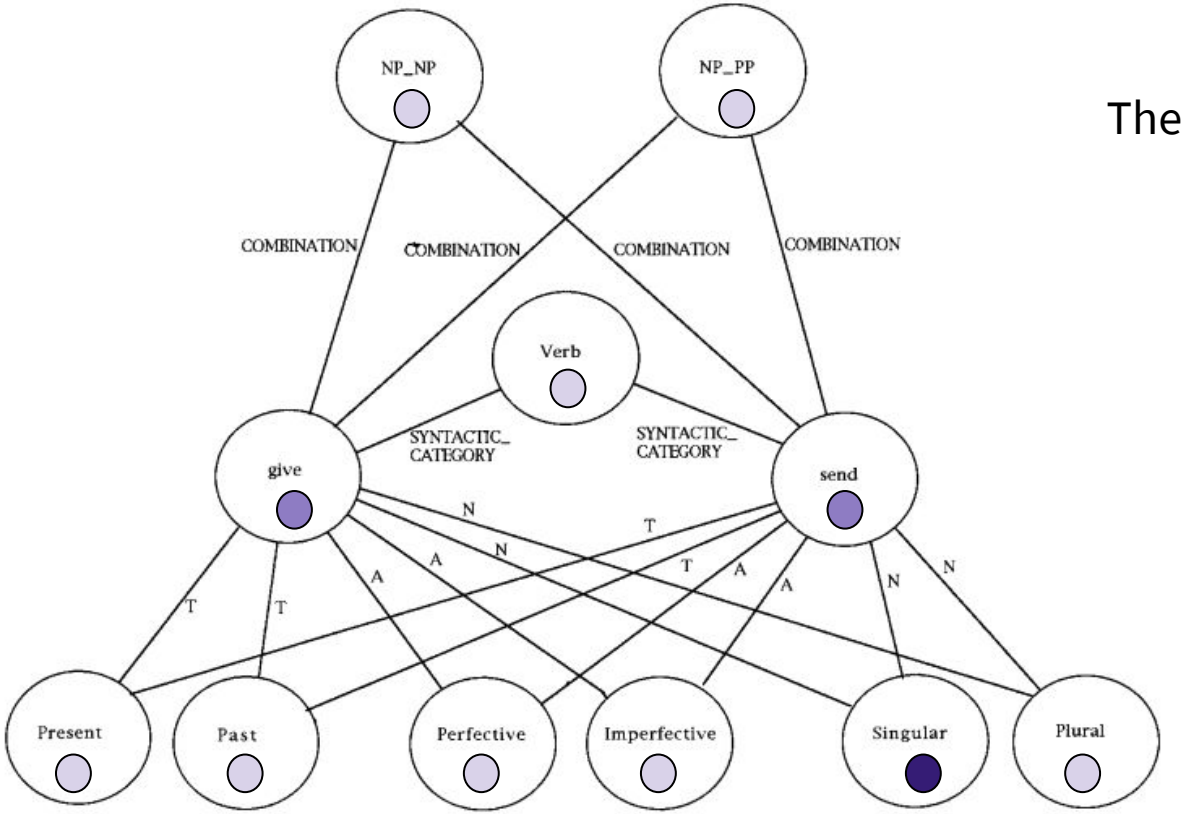
Pickering & Branigan (1998)



The girl …

Spreading activation

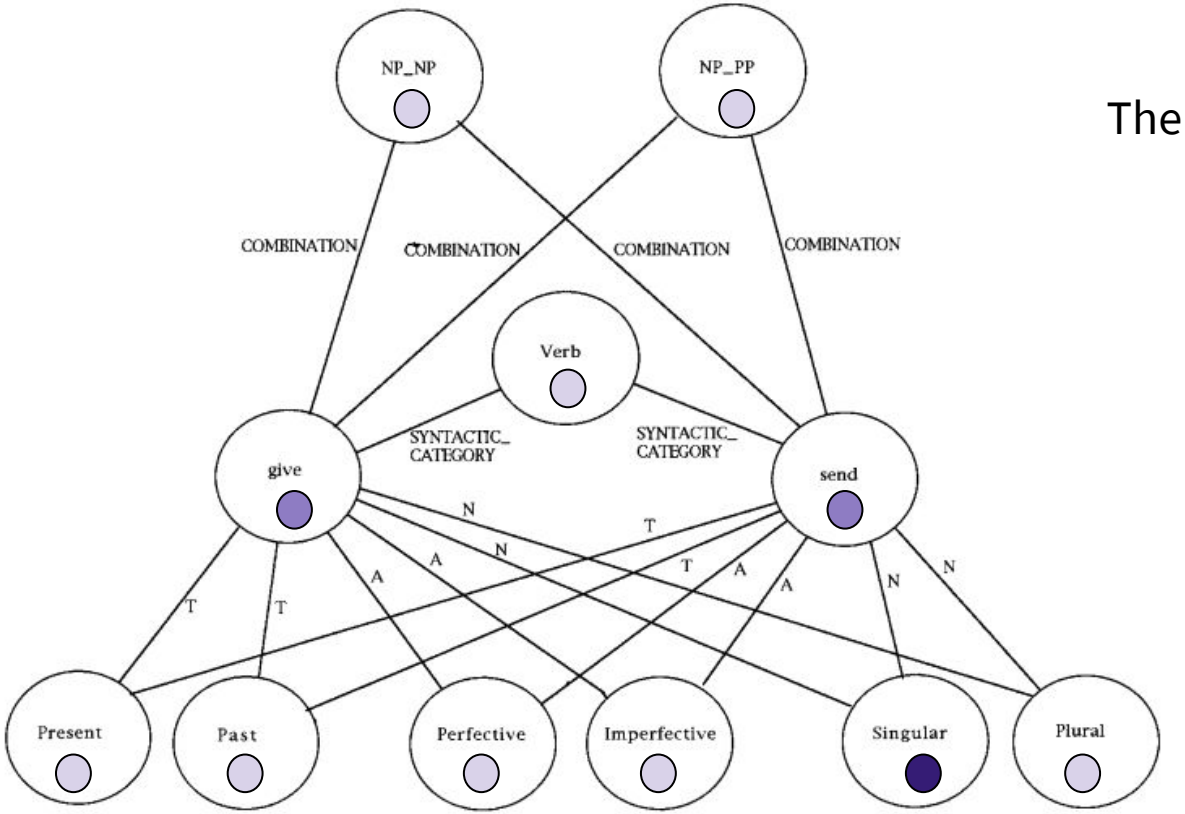
Pickering & Branigan (1998)



The girl …

Spreading activation

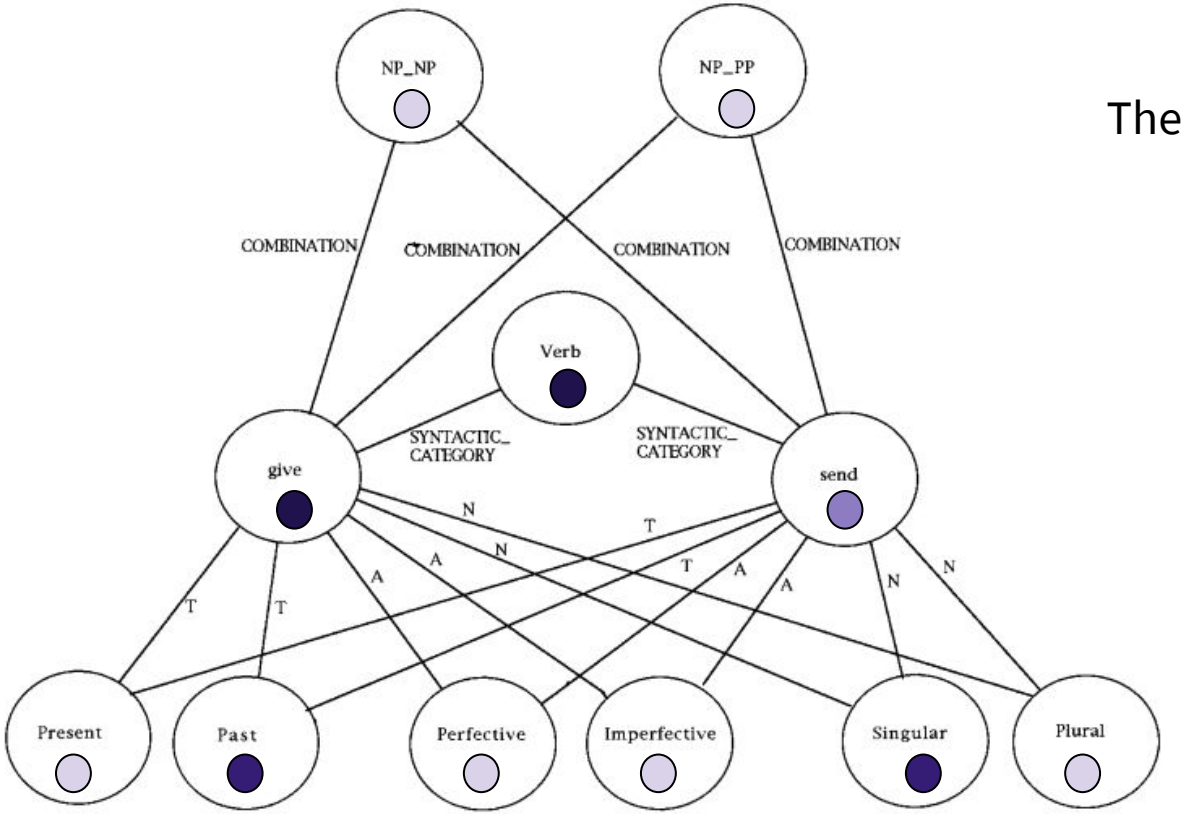
Pickering & Branigan (1998)



The girl gave

Spreading activation

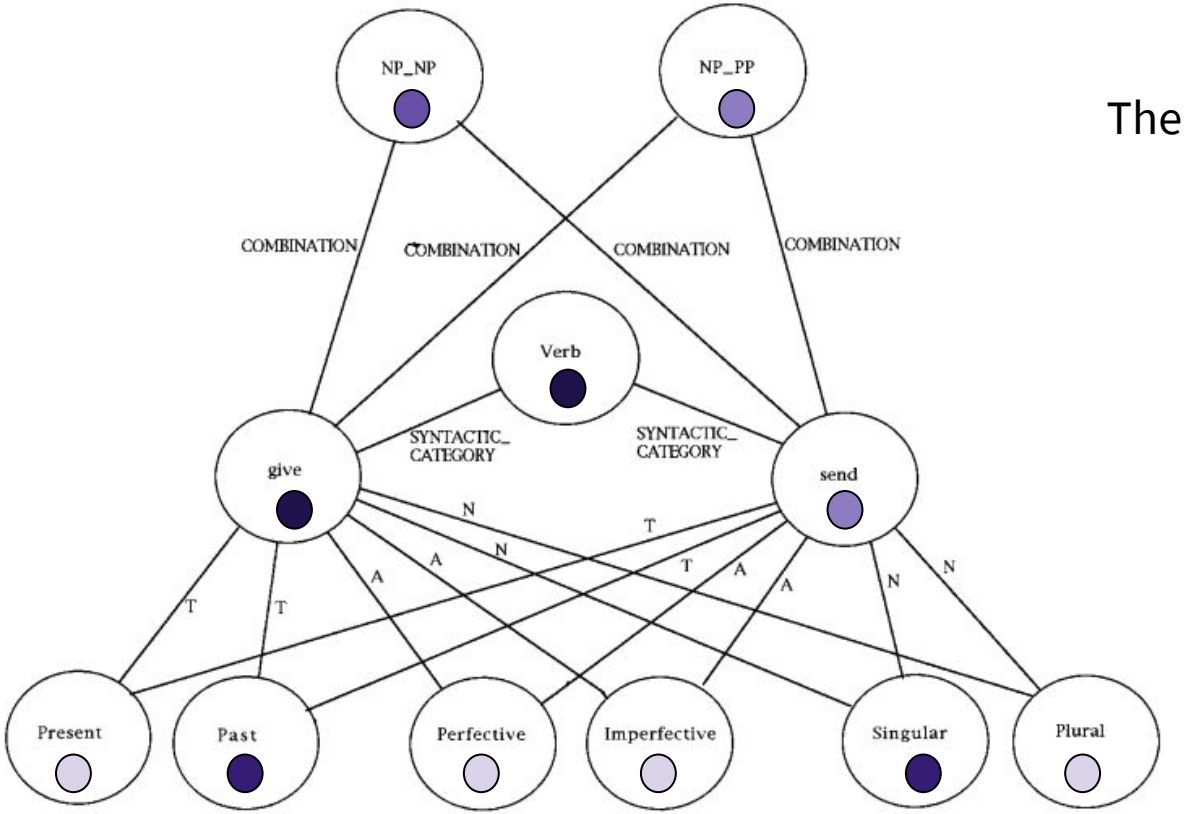
Pickering & Branigan (1998)



The girl gave

Spreading activation

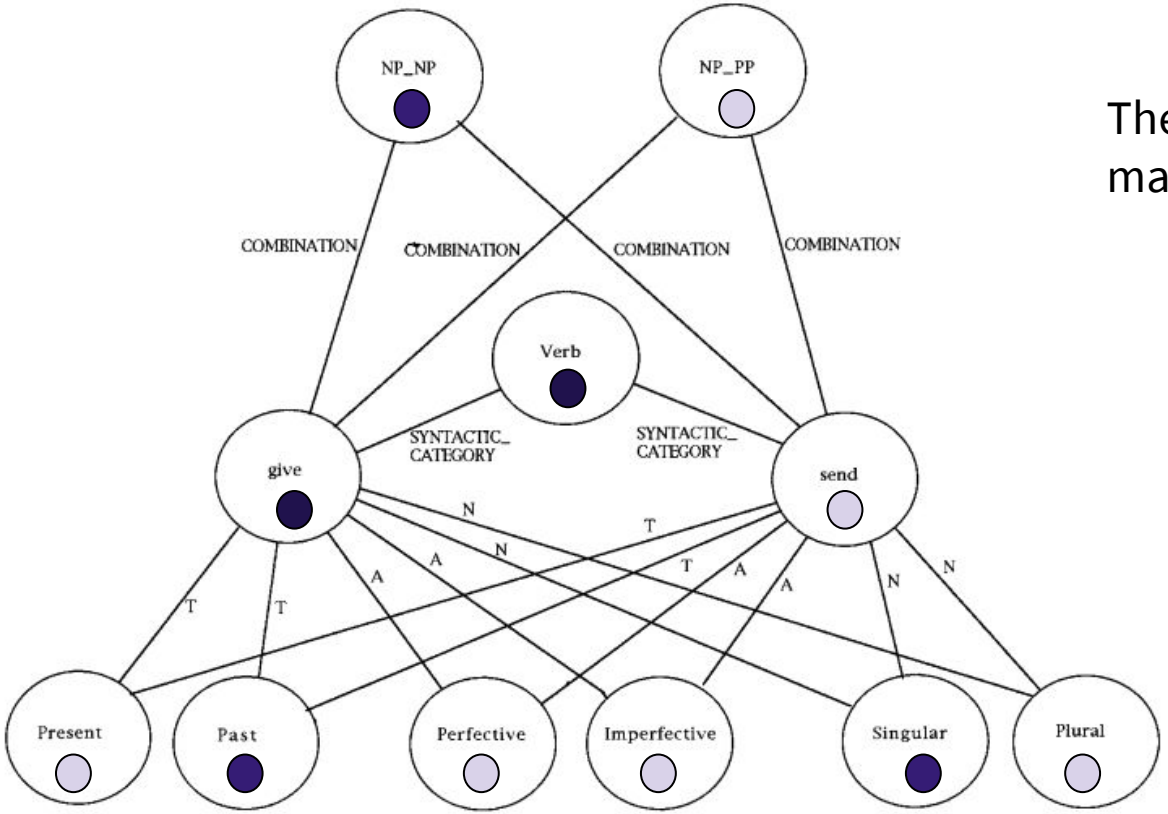
Pickering & Branigan (1998)



The girl gave

Spreading activation

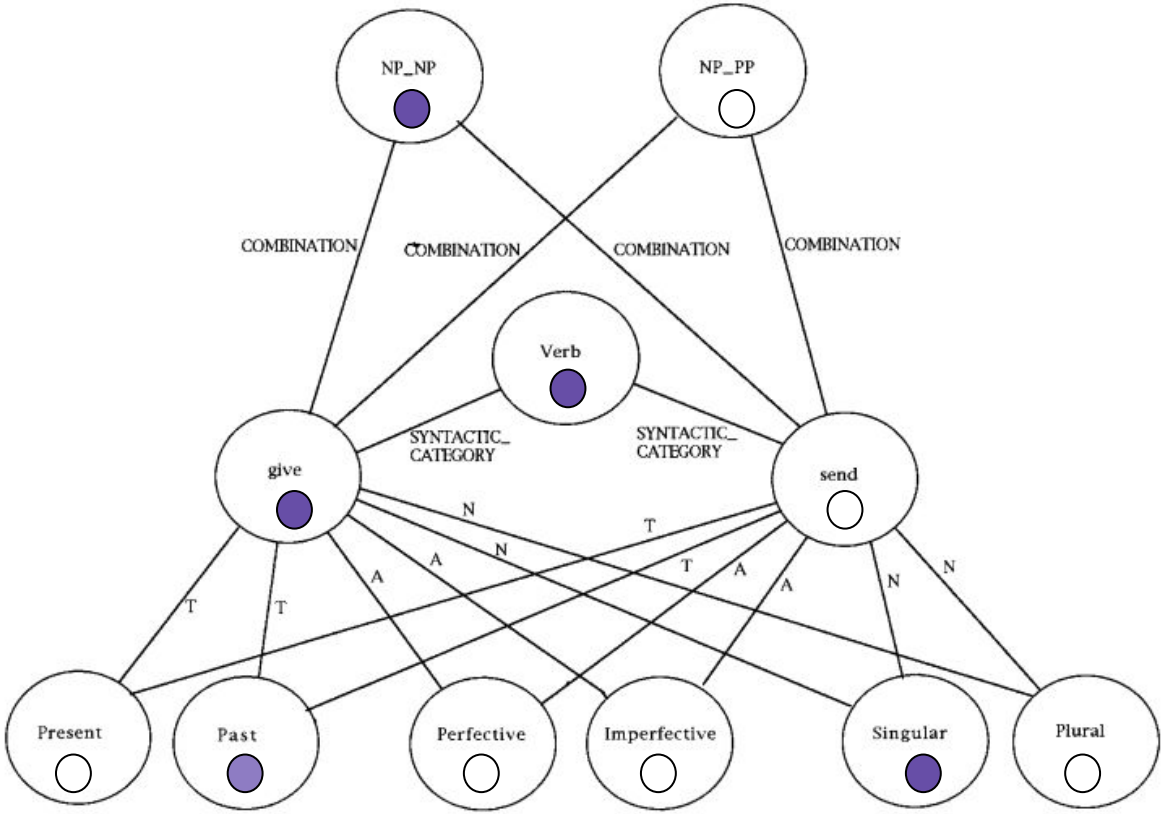
Pickering & Branigan (1998)



The girl gave the man the paintbrush

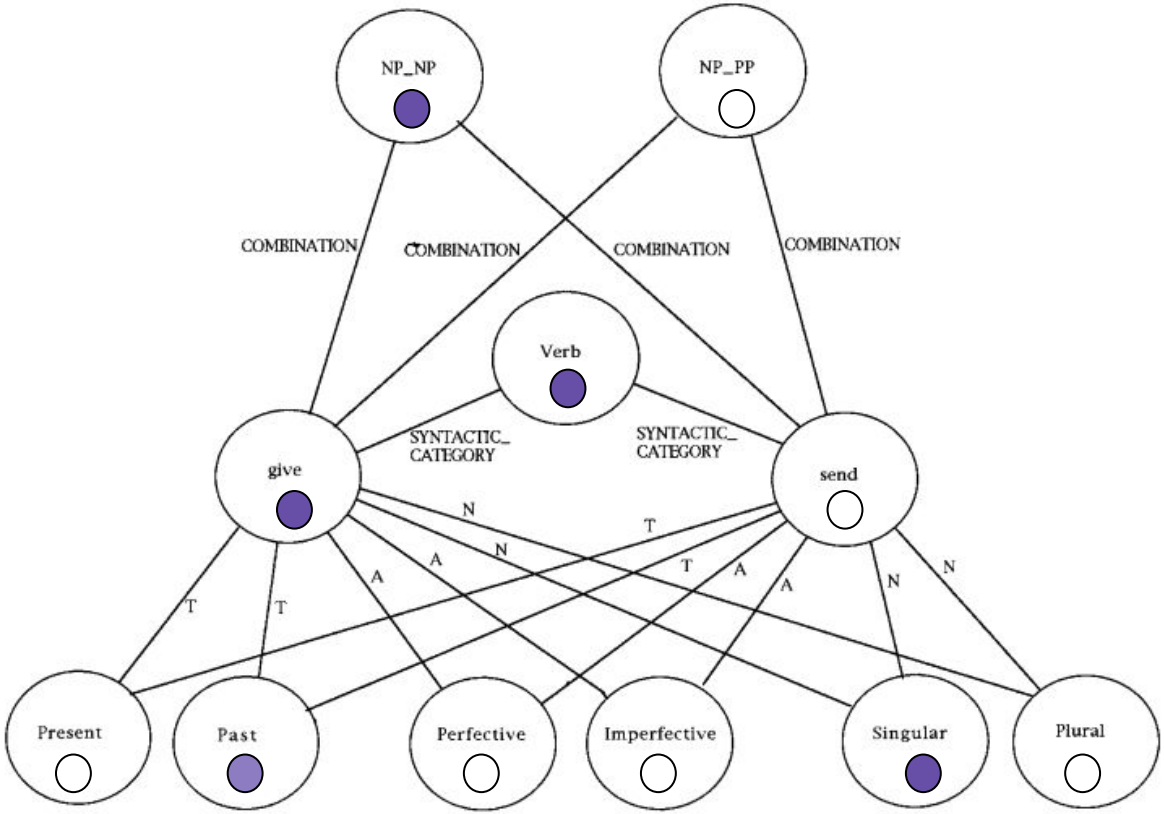
Spreading activation

Pickering & Branigan (1998)



Spreading activation

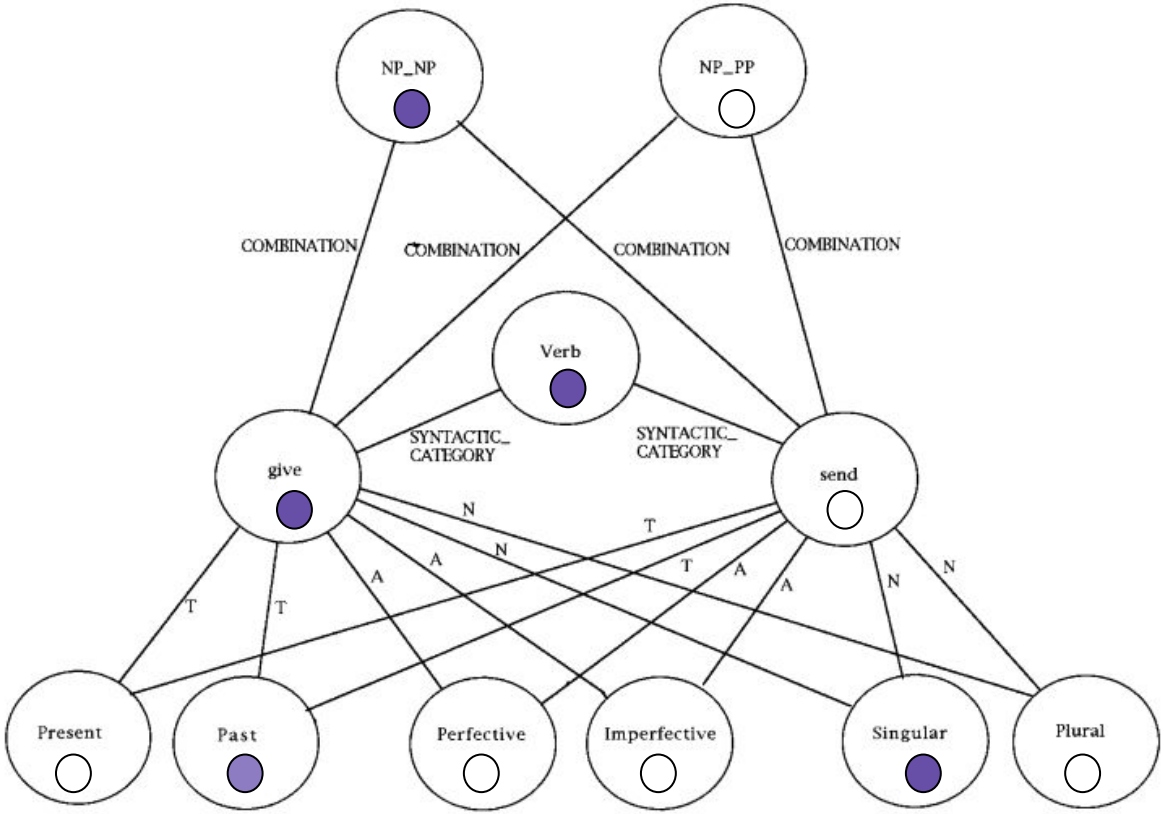
Pickering & Branigan (1998)



Residual activation makes it easier to process/ produce similar structures

Spreading activation

Pickering & Branigan (1998)



Residual activation makes it easier to process/ produce similar structures

What predictions does this make?

Modeling spreading activation

Dubey et al (2006)

- Goal: Model *syntactic parallelism* effect (Frazier et al, 2000) as a general case of syntactic priming
  - *Syntactic Parallelism:* If the structure to the right of a conjunction is the same as the left, it is easier to process.
  - Det Adj N **and** Det Adj N easier than Det N **and** Det Adj N
- Models
  - Copy model : Not priming
  - Within model : Within sentence priming
  - Between model : Between sentence priming

Modeling spreading activation

Dubey et al (2006)

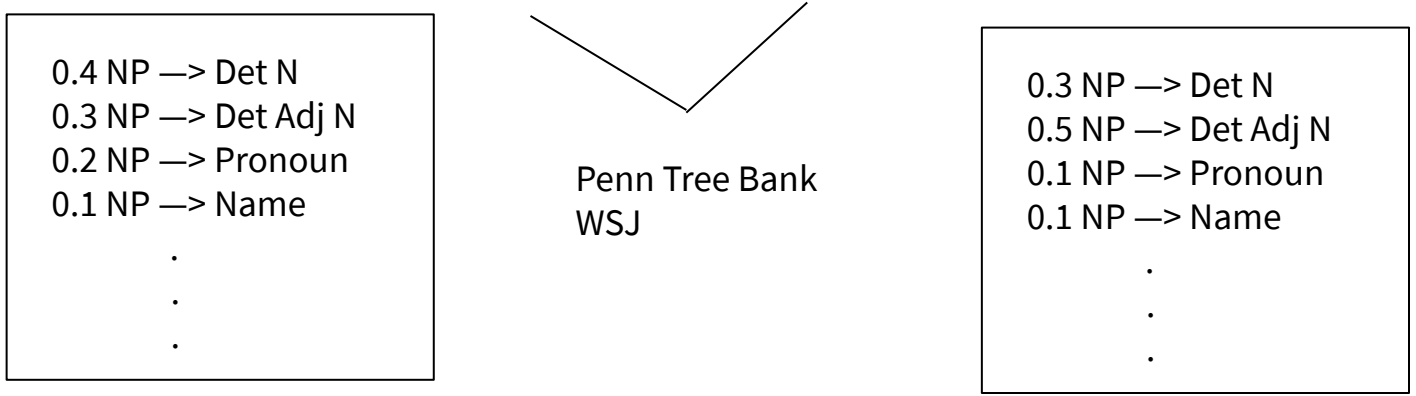
- Goal: Model *syntactic parallelism* effect (Frazier et al, 2000) as a general case of syntactic priming
  - *Syntactic Parallelism:* If the structure to the right of a conjunction is the same as the left, it is easier to process.
  - Det Adj N **and** Det Adj N easier than Det N **and** Det Adj N
- Models
  - Copy model : Not priming
  - Within model : Within sentence priming
  - Between model : Between sentence priming

Use spreading activation idea to model priming

Modeling spreading activation

- Basic idea: Use a cache to keep track of recently seen rules (i.e. rules with “residual activation”)
- There are two look up tables: One for when there is a prime (icurrent rule in the cache) and one for when there isn’t a prime (current rule not in the cache)

$$P(Det\ N|NP) = \frac{c(NP \rightarrow Det\ N)}{c(NP)} \qquad P(Det\ N|NP, Prime = 1) = \frac{c(NP \rightarrow Det\ N, Prime = 1)}{c(NP, Prime = 1)}$$



Modeling spreading activation

What counts as context?

- Between model:

Modeling spreading activation

What counts as context?

- Between model: Rules used in the previous sentence

Modeling spreading activation

What counts as context?

- Between model: Rules used in the previous sentence

What happens with globally ambiguous sentences?

Modeling spreading activation

What counts as context?

- Between model: Rules used in the previous sentence
- Within model:

What happens with globally ambiguous sentences?

Modeling spreading activation

What counts as context?

What happens with globally ambiguous sentences?

- Between model: Rules used in the previous sentence
- Within model: Rules used in the sentence so far
  - Counts once the parser passes word on the left most corner
    - E.g. The man in the blue shirt … NP —> Det N PP starts at **The**

Modeling spreading activation

What counts as context?

- Between model: Rules used in the previous sentence
- Within model: Rules used in the sentence so far
  - Counts once the parser passes word on the left most corner
    - E.g. The man in the blue shirt … NP —> Det N PP starts at **The**
  - Which parse to use? The used most probable or the correct parse

What happens with globally ambiguous sentences?

What else could they have done?

Modeling spreading activation

What counts as context?

- Between model: Rules used in the previous sentence
- Within model: Rules used in the sentence so far
  - Counts once the parser passes word on the left most corner
    - E.g. The man in the blue shirt … NP —> Det N PP starts at **The**
  - Which parse to use? The used most probable or the correct parse

What happens with globally ambiguous sentences?

What else is missing?

What else could they have done?

Modeling spreading activation

What counts as context?

- Between model: Rules used in the previous sentence
- Within model: Rules used in the sentence so far
  - Counts once the parser passes word on the left most corner
    - E.g. The man in the blue shirt … NP —> Det N PP starts at **The**
  - Which parse to use? The used most probable or the correct parse

What happens with globally ambiguous sentences?

What else is missing? **Decay! — model it with ACT-R in a later paper**

What else could they have done?

Implicit learning (a.k.a. cumulative priming)

People maintain probability distributions of structures and they update these distributions based on the input they receive

Implicit learning (a.k.a. cumulative priming)

People maintain probability distributions of structures and they update these distributions based on the input they receive

0.9	NP → Det N	(90/100)
0.1	NP → Det N RRC	(10/100)

[The soldiers warned about [the danger]$_{\text{Det N}}$]$_{\text{Det N RRC}}$ conducted [the raid]$_{\text{Det N}}$ x 40

Implicit learning (a.k.a. cumulative priming)

People maintain probability distributions of structures and they update these distributions based on the input they receive

0.77	~~0.9~~	NP → Det N	~~(90/100)~~ (90 + 80 / 100 + 80 + 40)
	0.1	NP → Det N RRC	(10/100)

[The soldiers warned about [the danger]$_{\textbf{Det N}}$]$_{\text{Det N RRC}}$ conducted [the raid]$_{\textbf{Det N}}$ x **40**

Implicit learning (a.k.a. cumulative priming)

People maintain probability distributions of structures and they update these distributions based on the input they receive

0.77	~~0.9~~	NP → Det N	~~(90/100)~~ (90 + 80 / 100 + 80 + 40)
0.33	~~0.1~~	NP → Det N RRC	~~(10/100)~~ (10 + 40 / 100 + 80 + 40)

[The soldiers warned about [the danger]$_{\text{Det N}}$]$_{\textbf{Det N RRC}}$ conducted [the raid]$_{\text{Det N}}$ x **40**

Implicit learning (a.k.a. cumulative priming)

People maintain probability distributions of structures and they update these distributions based on the input they receive

0.77	~~0.9~~	NP → Det N	~~(90/100)~~ (90 + 80 / 100 + 80 + 40)
0.33	~~0.1~~	NP → Det N RRC	~~(10/100)~~ (10 + 40 / 100 + 80 + 40)

More fancy ways to update beliefs!

[The soldiers warned about [the danger]$_{\text{Det N}}$]$_{\textbf{Det N RRC}}$ conducted [the raid]$_{\text{Det N}}$ x **40**

Implicit learning (a.k.a. cumulative priming)

People maintain probability distributions of structures and they update these distributions based on the input they receive

0.77	~~0.9~~	NP → Det N	~~(90/100)~~ (90 + 80 / 100 + 80 + 40)
0.33	~~0.1~~	NP → Det N RRC	~~(10/100)~~ (10 + 40 / 100 + 80 + 40)

More fancy ways to update beliefs!

[The soldiers warned about [the danger]$_{\text{Det N}}$]$_{\textbf{Det N RRC}}$ conducted [the raid]$_{\text{Det N}}$ x **40**

How do the predictions made by the implicit learning account differ from those made by spreading activation account?

Implicit learning (a.k.a. cumulative priming)

People maintain probability distributions of structures and they update these distributions based on the input they receive

0.77	~~0.9~~	NP → Det N	~~(90/100)~~ (90 + 80 / 100 + 80 + 40)
0.33	~~0.1~~	NP → Det N RRC	~~(10/100)~~ (10 + 40 / 100 + 80 + 40)

More fancy ways to update beliefs!

[The soldiers warned about [the danger]$_{\text{Det N}}$]$_{\textbf{Det N RRC}}$ conducted [the raid]$_{\text{Det N}}$ x **40**

How do the predictions made by the implicit learning account differ from those made by spreading activation account?

You can have priming effects across multiple sentences over long periods of time

Modeling cumulative priming

van Schijndel & Linzen (2018) — Adaptive neural model

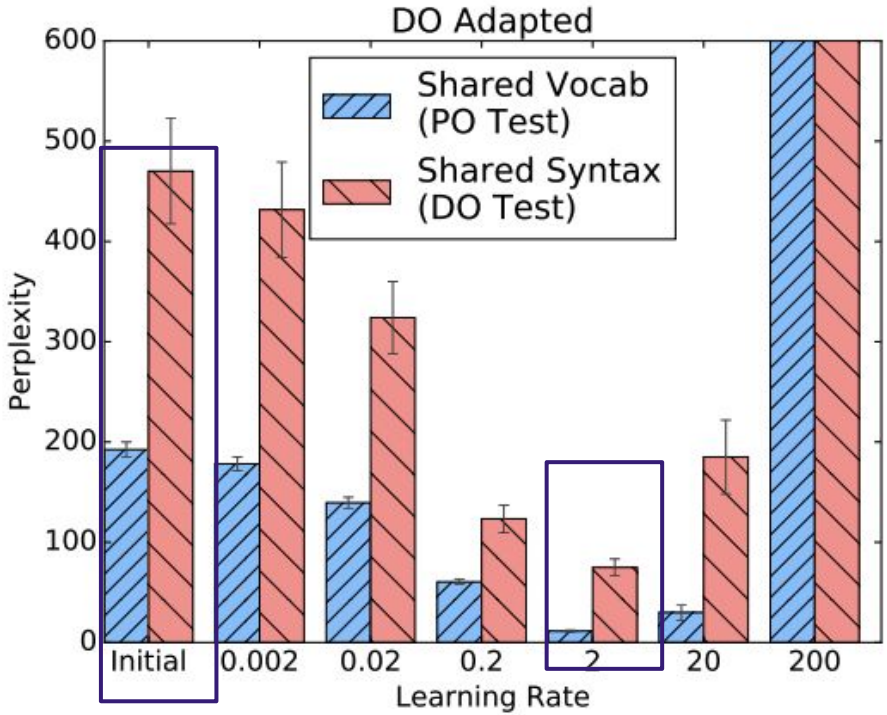
Modeling cumulative priming

van Schijndel & Linzen (2018) — Adaptive neural model

How does this relate to cumulative priming? What evidence would we need to know that this can model syntactic priming?

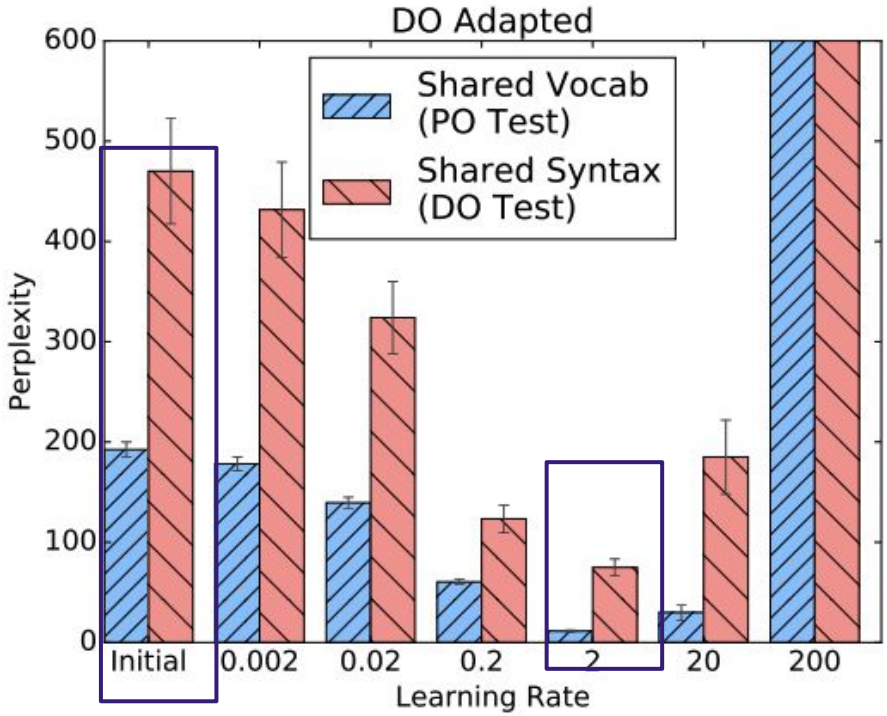
Modeling cumulative priming

van Schijndel & Linzen (2018) — Adaptive neural model



Modeling cumulative priming

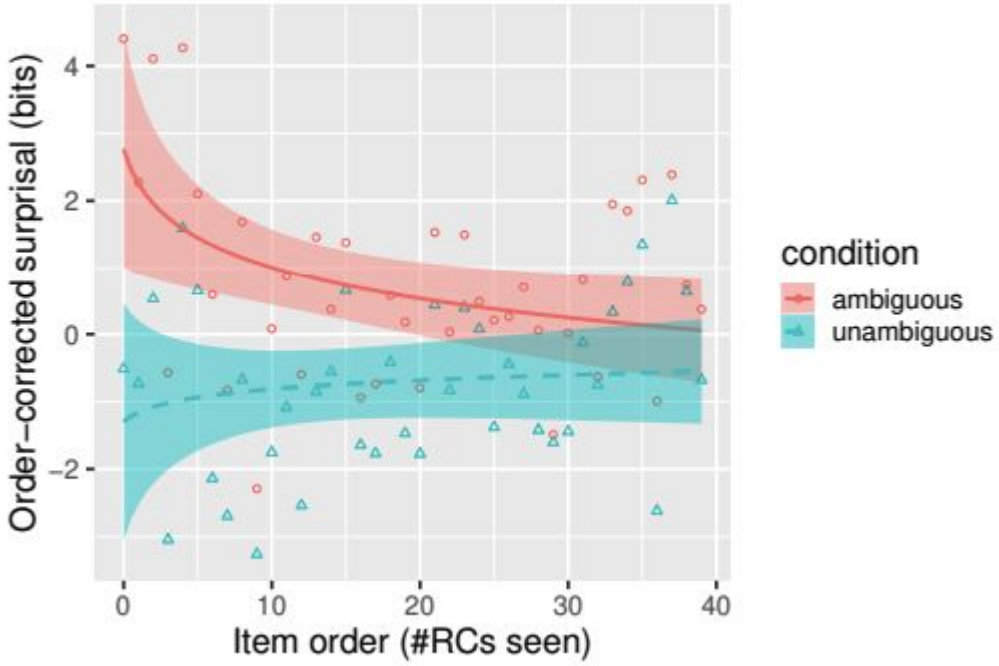
van Schijndel & Linzen (2018) — Adaptive neural model



When adapted to DO sentences, the perplexity (and surprisal) “processing” new DO sentences becomes “easier” (i.e. there is a decrease in surprisal or perplexity)

Modeling cumulative priming

van Schijndel & Linzen (2018) — Adaptive neural model



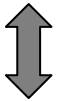
Decrease in garden path effect over time

Outline

1. What is syntactic priming?
   a. Production
   b. Comprehension
2. Why does syntactic priming occur?
   a. Spreading activation.
      - Computational implementation: Dubey et al (2006)
   b. Implicit learning
      - Computational implementation (ish): van Schijndel & Linzen (2018)
3. Graded syntactic priming and how it can provide insights into how neural language models (e.g. RNNs) and humans represent syntax.

But what does syntactic priming tell us?

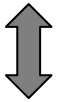
A stimulus is easier to produce/ process if it is preceded by a **related** stimulus.



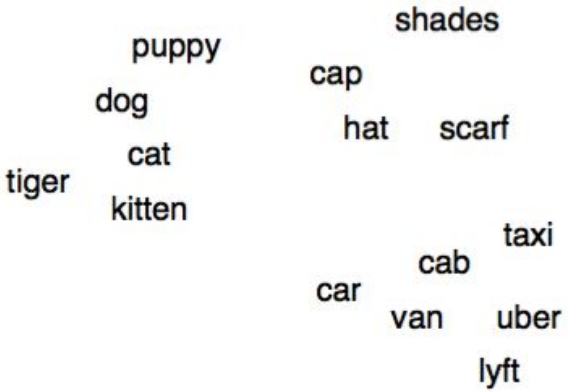
If stimulus A is easier to produce/process after stimulus B, then A and B are related

But what does syntactic priming tell us?

A stimulus is easier to produce/ process if it is preceded by a **related** stimulus.



If stimulus A is easier to produce/process after stimulus B, then A and B are related

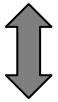


Dog primes cat more than car

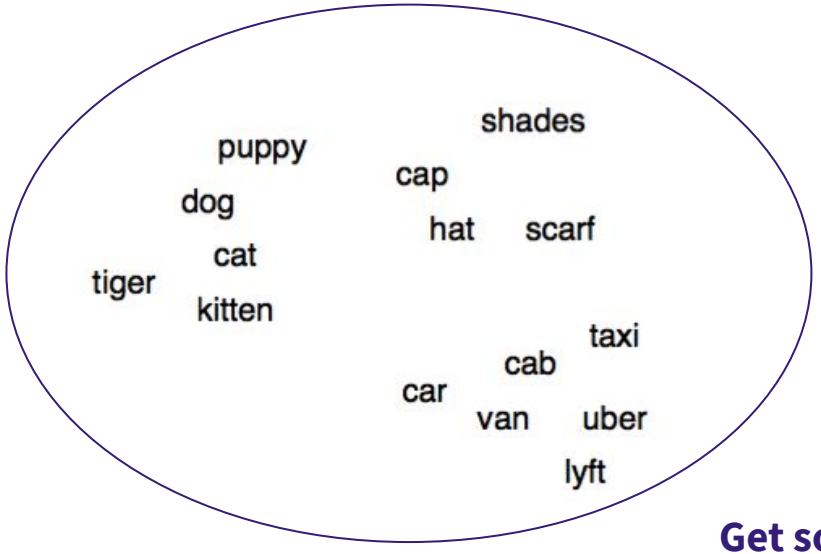
Dog is closer in the representation space to cat than to car

But what does syntactic priming tell us?

A stimulus is easier to produce/ process if it is preceded by a **related** stimulus.



If stimulus A is easier to produce/process after stimulus B, then A and B are related



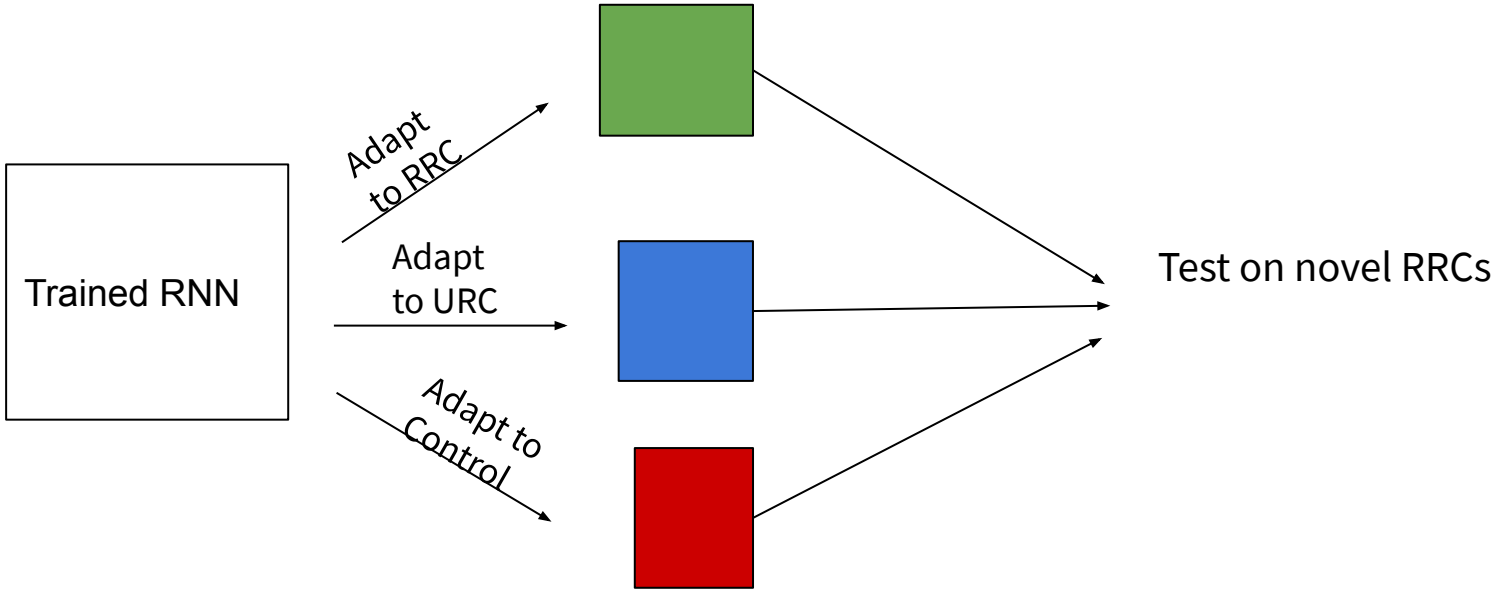
Dog primes cat more than car

Dog is closer in the representation space to cat than to car

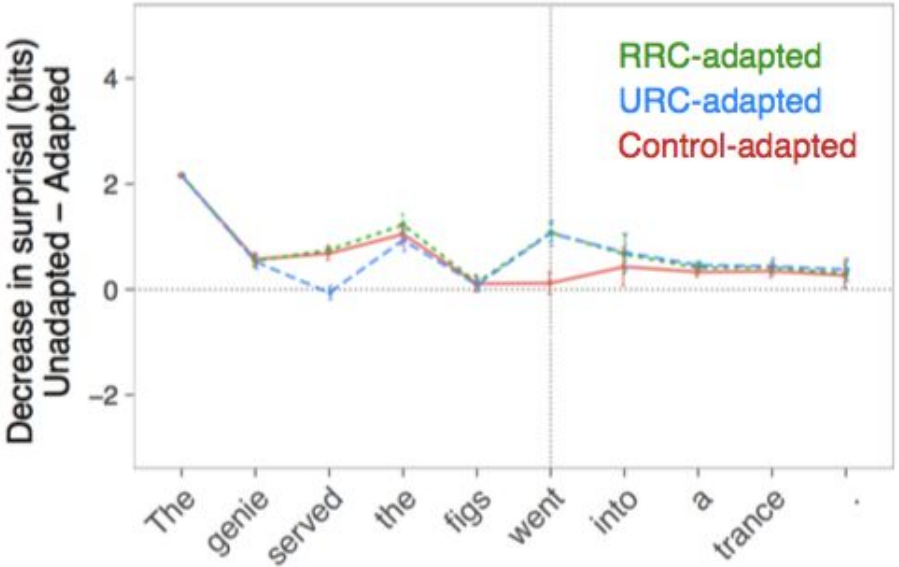
Get something like this but for syntactic structures

Proof of concept experiment

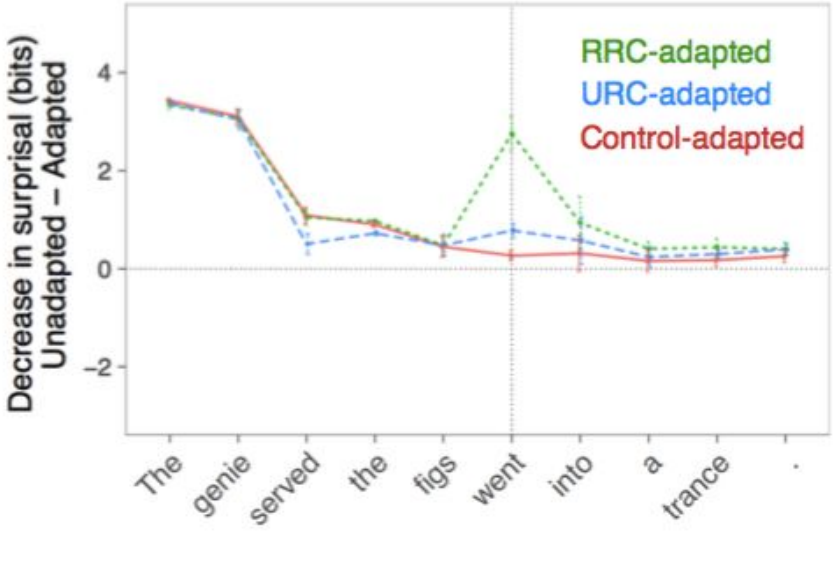
RRC: The six volunteers taught the complicated procedure **learned** it very well.
URC: The six volunteers who were taught the complicated procedure **learned** it very well.
Control: The six volunteers taught the complicated procedure and **learned** it very well.



Proof of concept experiment



2 Million words of Wiki. 200 hidden units



90 Million words of Wiki. 650 hidden units