Exemplars and constructions in syntactic production

Neal Snider

LSA Annual Meeting

January 4, 2008

Big question

- How is linguistic knowledge represented and stored?

- Model: Exemplars
- Evidence: Structural priming

Exemplar Models

- Linguistic knowledge consists of stored memory traces of finely grained linguistic experience
- New tokens are produced and comprehended based on similarity to previous exemplars

Hypothesis

- What do exemplar models predict about language processing?
- No explicit distinction between levels of word and structure

→Lexical access and structural access should be the same process

  - Evidence: priming

Outline

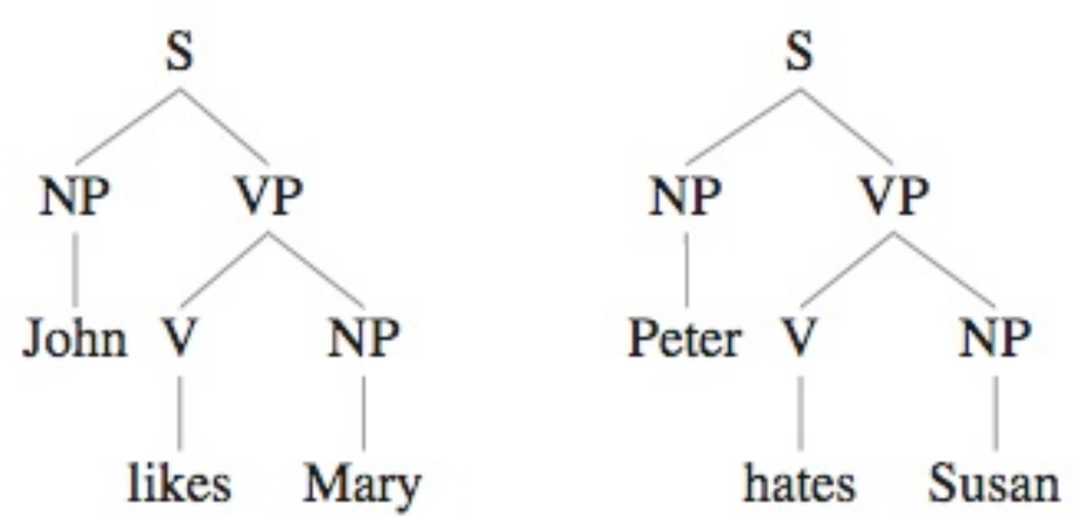
- Exemplars in syntax
- Lexical priming results
- Spreading activation model and predictions for structural priming
- Data - corpus studies of voice alternation
- Results
  - Prime frequency
  - Prime-target similarity
  - Prime neighborhood density

Exemplars in syntax

- Data Oriented Parsing (Bod 1992, 1998, 2006)
- Each sentence in the training corpus is broken into all possible subtrees
- New sentences are parsed by recombining subtrees (according to their frequency)

TreeDOP: Representation

- Example corpus:

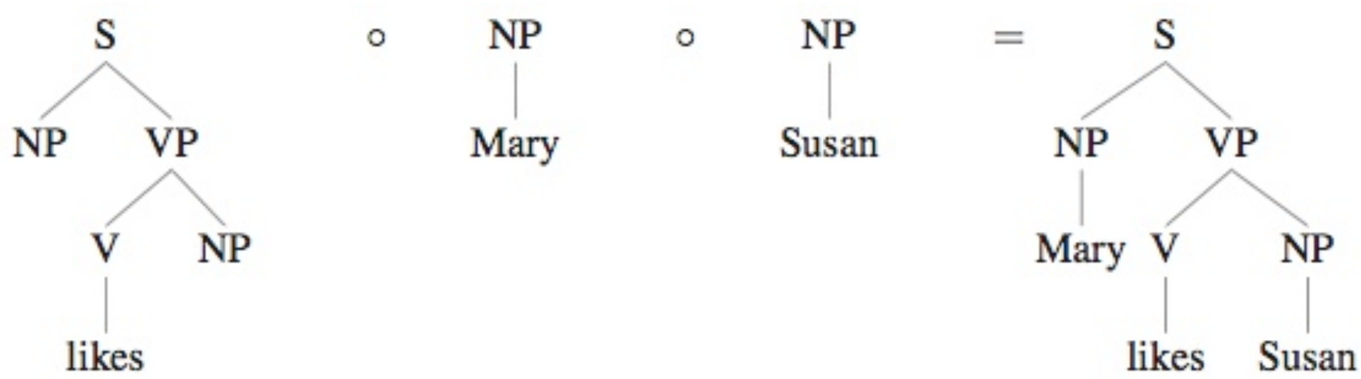


- Decomposition: this is not only a corpus of 2 S's, but also 4 NPs, 2 VPs, etc.

DOP: Corpus of subtrees

TreeDOP: Parsing

- How DOP parses *Mary likes Susan*:



- Exemplars have abstraction, like constructions
- DOP models for many types of representation (HPSG, LFG), even unsupervised!

Lexical priming

Lexical priming results

- Prime frequency
  - Low frequency words give more of a boost (Scarborough et al 1977; Versace and Nevers 2003; etc)
- Prime-target similarity
  - More similar primes give more of a boost (Ratcliff and McKoon 1981; 1988; Schreuder et al 1983)
- Prime neighborhood density
  - Primes that are similar to fewer words give more of a boost (Perea and Rosa 2000; Anaki and Henik 2003)

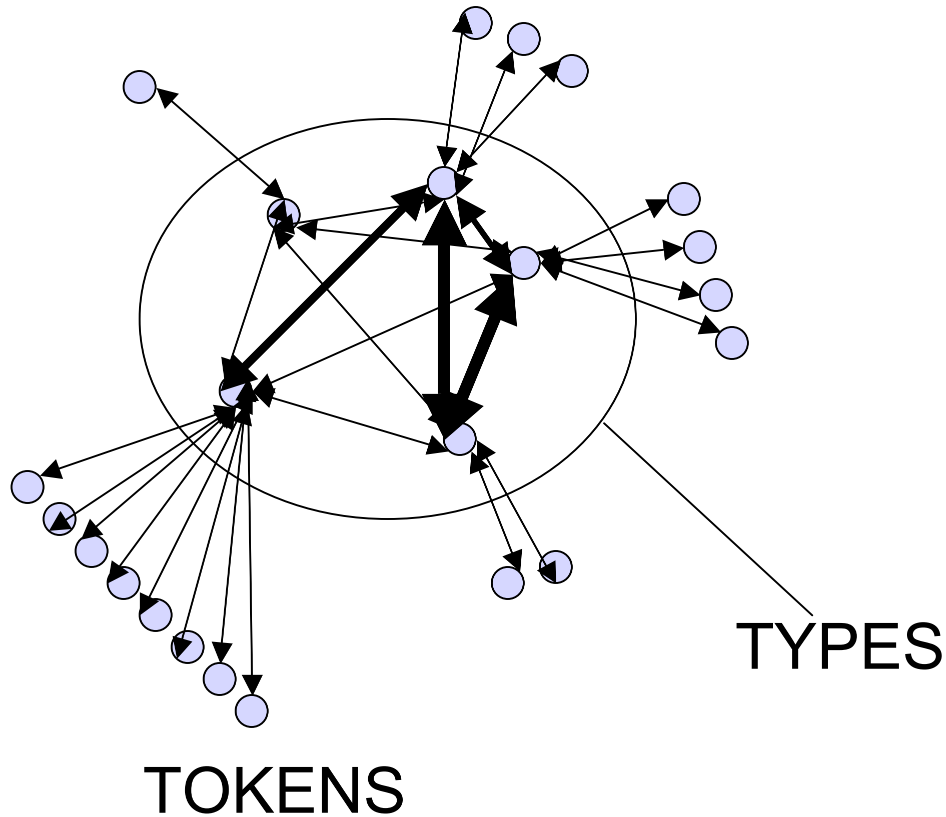
Spreading activation model

Spreading activation models

- Lexical access modeled as the spread of activation through a network of nodes
  - Nodes correspond to linguistic entities (words, phones, etc.)
- Used with success in phonology and morphology (McClelland and Rumelhart 1981, Anderson 1983)

LAST model (Kapatsinski 2006)

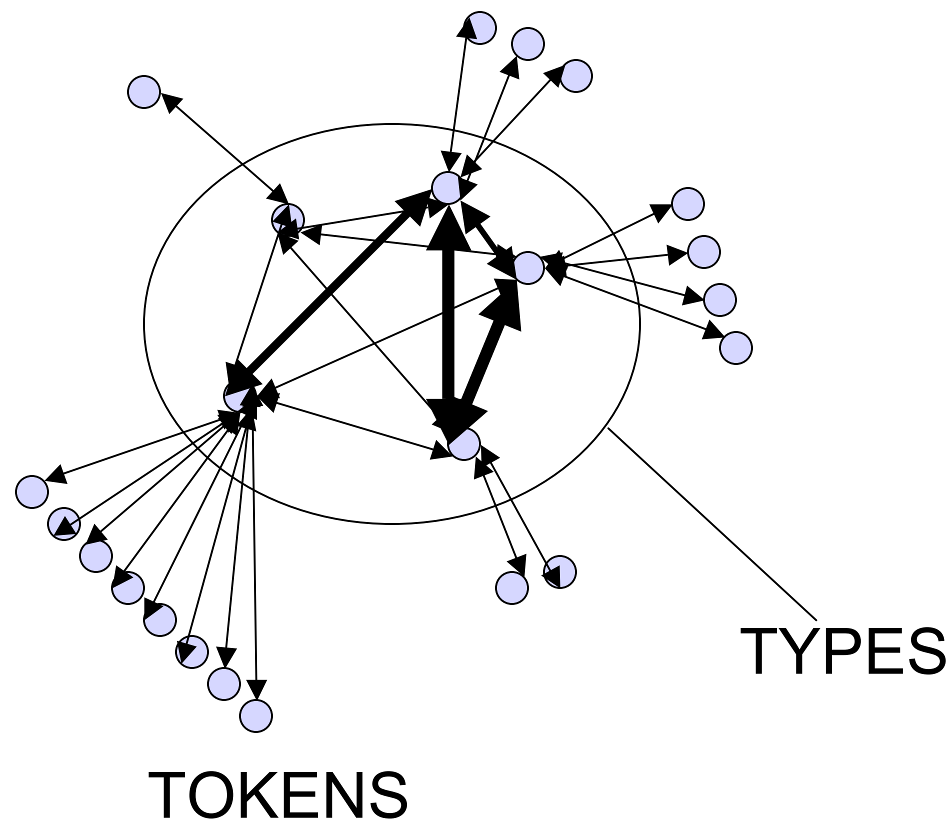
- Local Activation Spread Theory
- Memory is a network
- In this network, each unit corresponds to a node



LAST model (Kapatsinski 2006)

- There are type nodes and token nodes
- Every memorized chunk, e.g. a word, a morpheme, a phoneme, a construction, owns a type node
- Every presentation of a chunk forms a token node

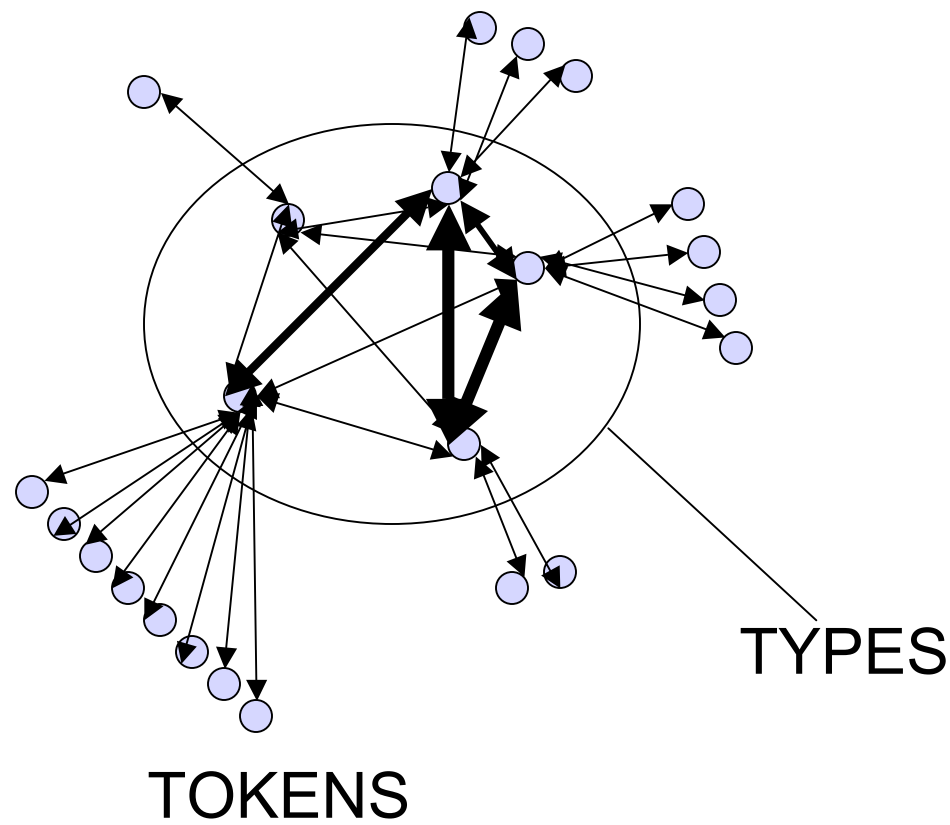
- Similar to construction grammar type hierarchy



LAST model (Kapatsinski 2006)

- Most or all of the token's activation spreads to one type (its best match)
- Every type is connected to all other types but connection strengths vary based on co-occurrence

- Predicts lexical priming effects of frequency, similarity, and neighborhood density



DOP-LAST model

- Each subtree in the corpus is a token of a particular type
- Types can be connected in a network based on co-occurrence
- In DOP, words and structures have same status in storage

- What does this predict for structural priming?

Structural priming in dialogue

- Tendency for syntactic structures to persist in corpora: (Sankoff and Laberge 1978; Poplack 1980; Weiner and Labov 1983; Szmrecsanyi 2005; Gries 2005)

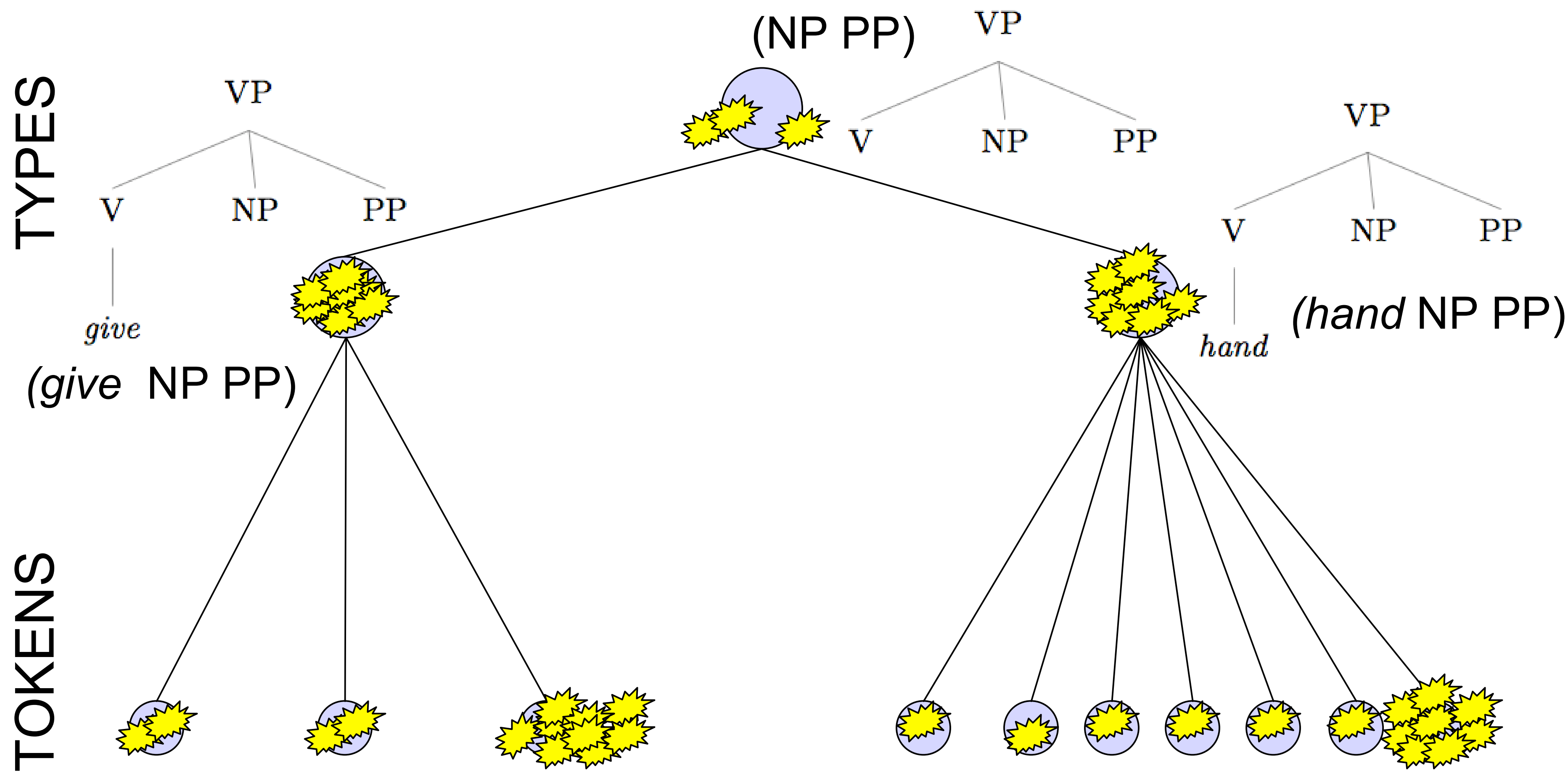
“… I don't feel we should **loan [them] [money].** … I wish our leaders were really seeking the Lord on these things, and if we feel led to **give [a country] [money]** to help them, fine”
(Switchboard corpus)

Predictions for structural priming

- Prime construction is more likely to be repeated when:
  - Infrequent
  - Target is structurally and semantically similar
  - In a low-density neighborhood

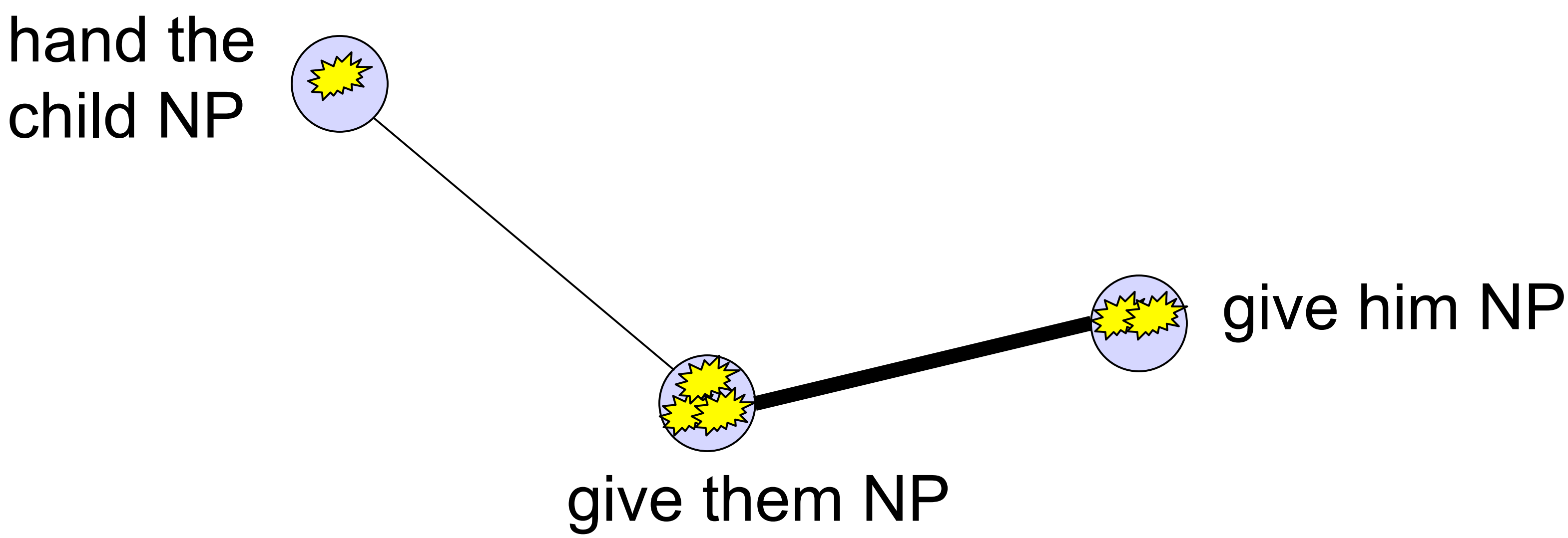
- How does DOP-LAST predict this?

Prime token frequency



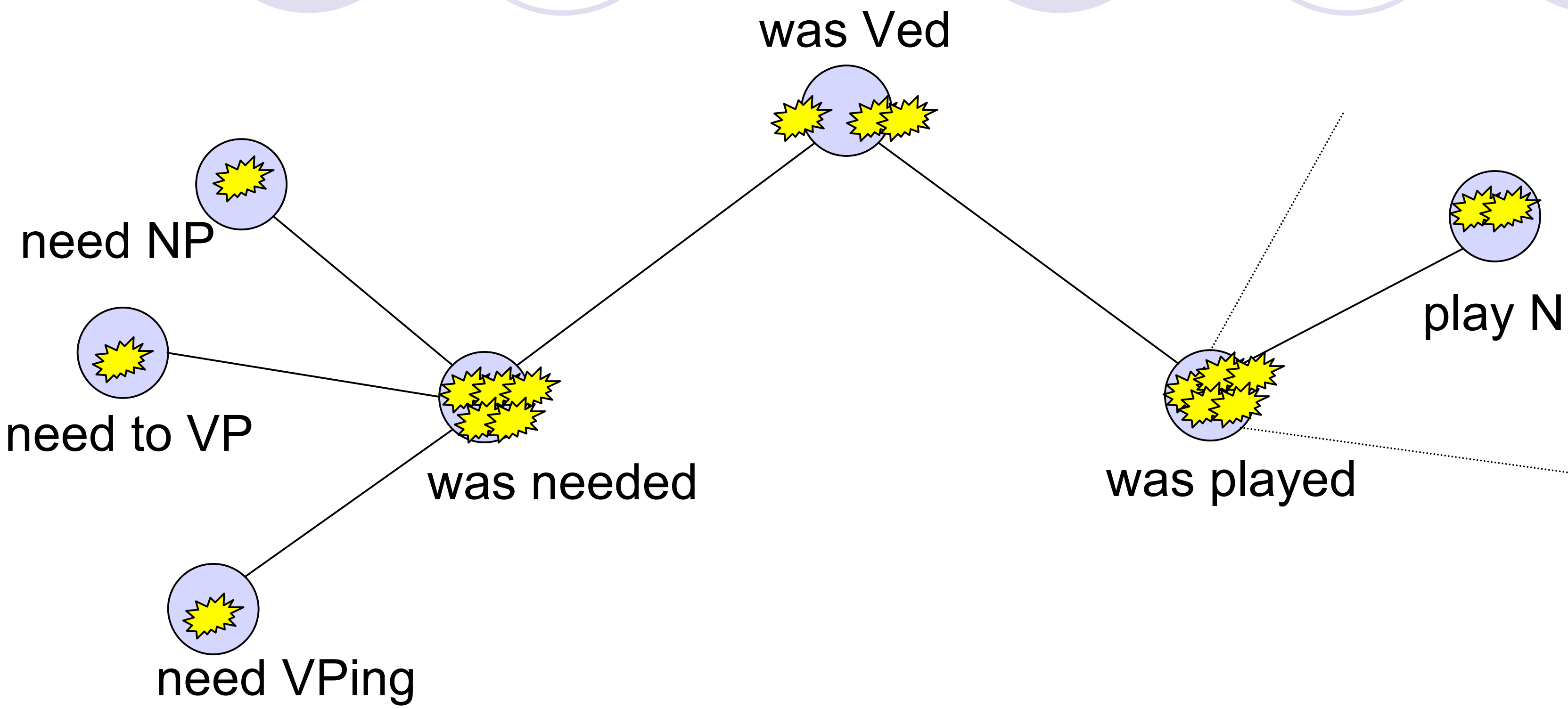
Prime types with more tokens leave less activation

Prime similarity



More activation flows into more similar types

Prime neighborhood density



Prime verbs that occur in more constructions leave less activation for the construction in which they occur

Summary so far

- Exemplar models predict words and structures are stored in the same structure
- Therefore, structural priming should behave like lexical priming

- This will provide first psycholinguistic evidence for exemplars in syntax

Data

Data

- Modeling syntactic choice in the voice alternation:

Passive	the thief was arrested
Active (impersonal)	they arrested the thief

- Following Weiner and Labov (1983) the true syntactic variable in the voice alternation is passive and impersonal active

Data

- 1757 actives and 557 passives from Switchboard corpus (all primed turns)
- Naturalistic, spontaneous speech, conversational data
- Analyzed with mixed-model logistic regression
  - positive response is Passive

Controls

- **Patient**
  - *Pronominality*
  - *Givenness*
  - *Definiteness*
  - *Log length*
  - *Animacy*
- **Speaker**
  - *Sex*
  - *Age*
- **Verb**
  - *Target verb bias*
- **Also modeled random effect of speaker**

Result 1: Passive prime frequency

Exemplar frequency

- In this study, exemplar frequency was taken to be the verb bias:

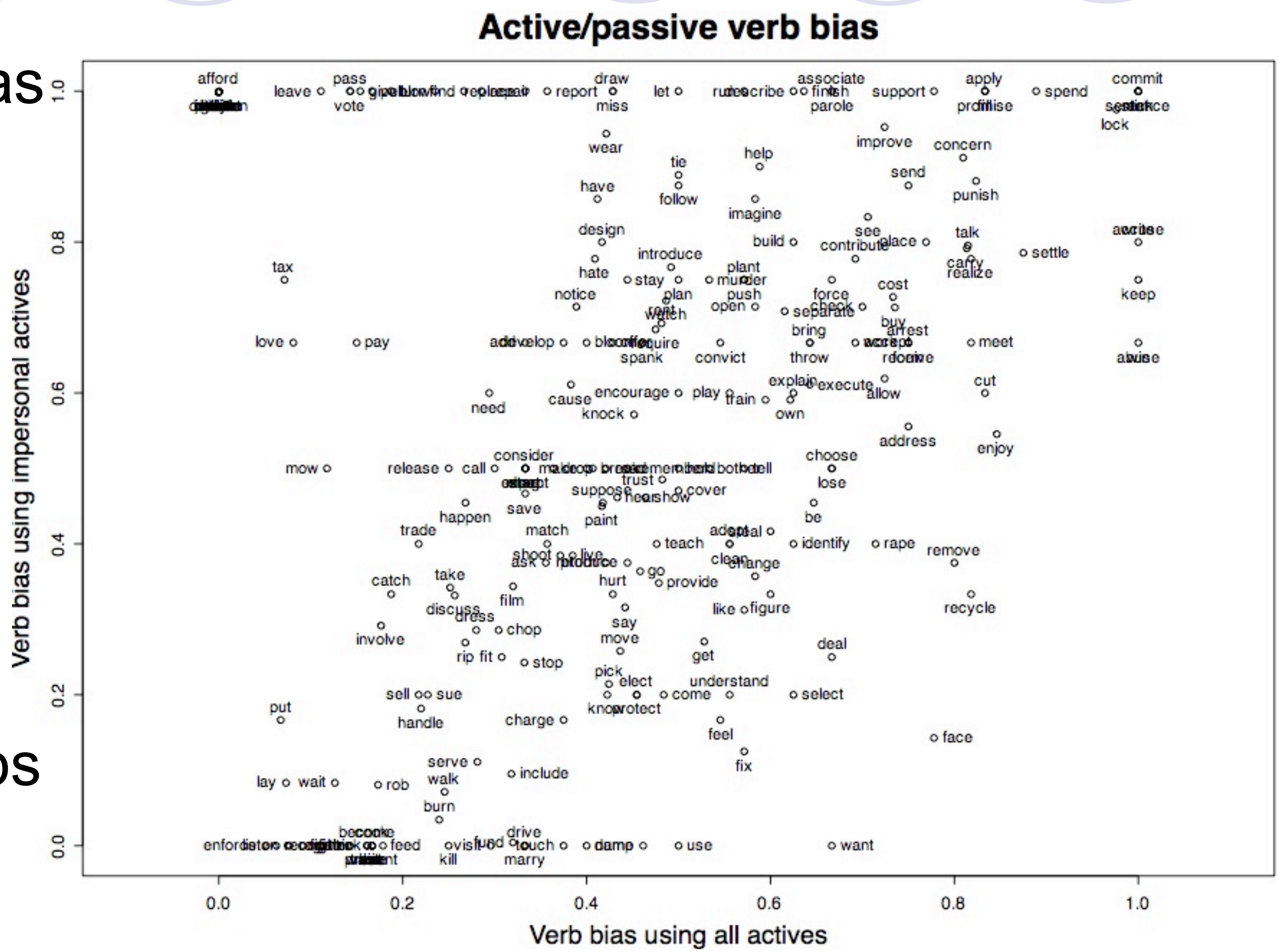
  P(Passive|verb)

  - estimated from Switchboard data set

- Because of inverse frequency effect in priming, verb bias is expected to have a **negative** effect on priming

Passive biased verbs?

- Used bias given all actives



- 219 verbs

Results of interest

Predictor (indpendent variable)	*Parameter estimates* Log-odds	S.E.	Odds	*Wald's test* *p*	$\Delta_\chi(\Lambda)$*-test* *p*
Prime Bias x Prime	-0.92	0.41	0.39	<.03	<.05
Target Bias	3.76	0.40	42.9	<<.001	<.001

- **If prime is Passive and verb is Active-biased → Passive more likely**
- Controls included, so priming effects not due to prime and target being in similar environments

Summary so far

- Just as in lexical priming, structural priming is sensitive to prime frequency
- In passive, less frequent prime constructions more likely to be repeated
  - Experimentally reproduced (Bernolet and Hartsuiker 2007)
- As predicted by exemplar hypothesis and DOP-LAST
  - Also demonstrates that **token** storage in the model is essential

Result 2: Similarity in passive priming

Priming and similarity

- As with lexical priming:
  - More similar exemplars should be more likely to prime!
- How to measure similarity?
- Nearest-neighbor models offer precise metric of exemplar similarity

k-NN similarity metric

- Distance is weighted sum of differences between all the features: (Daelemans 2005)

$$\Delta(X,Y) = \sum_i w_i \delta(x_i, y_i)$$

- Where the difference is defined as:

$$\delta(x_i, y_i) = \begin{cases} \frac{x_i - y_i}{max_i - min_i} & \text{if numeric, otherwise} \\ 0 & \text{if } x_i = y_i \\ 1 & \text{if } x_i \neq y_i \end{cases}$$

Features

- **Patient**
  - *Pronominality*
  - *Givenness*
  - *Definiteness*
  - *Log length*
  - *Animacy*
  - *Head word*

- **Verb**

Question

- Are exemplars that have less k-NN distance more likely to prime?
- BUT, is that effect merely due to known verb repetition boost? (Pickering and Branigan 1998)

Results

Predictor (indpendent variable)	*Parameter estimates* Log-odds	S.E.	Odds	*Wald's test* p	$\Delta_{\chi}(\Lambda)$*-test* p
NN distance x Prime	-3.3	1.23	0.72	<.008	<.03
primeV=targV x Prime	-0.17	1.16	0.84	<.9	<.2

- Controls replicate
- Effect is not due to known verb repetition boost

Summary so far

- Just as in lexical priming:
- Structural priming is affected by
  - (Inverse) exemplar frequency
  - Exemplar similarity

- Evidence that lexical production and structural production access same representations
  - As predicted by exemplar model

Result 3: Prime neighborhood density

Prime neighborhood density

- Prime verbs that occur in many constructions are in dense neighborhoods
- Extract all 14 constructions from Roland et al (2007) in Switchboard corpus:
  - [V], [V PP], [V NP], [V NP PP], [V NP NP], [V to VP], [V VP], [V VPing], etc.
  - High density verb: *need* (12)
  - Low density verb: *play* (1)

Possible confound

- Verbs that occur in many constructions are also frequent
- Control by modeling prime verb frequency

Results

Predictor (indpendent variable)	*Parameter estimates* Log-odds	S.E.	Odds	*Wald's test* p	$\Delta_{\chi}(\Lambda)$*-test* p
Prime density x Prime	-0.62	.25	0.54	<.02	<.03
Prime V freq x Prime	-0.01	.99	42.9	<.25	<.2

- Controls replicate
- Prime constructions in dense neighborhoods are less likely to be repeated

Summary so far

- Constructions in dense neighborhoods are less likely to prime
  - Similar to lexical priming
  - As predicted by exemplar hypothesis and DOP-LAST model
- Demonstrates necessity of fully-connected **types**
  - abstraction is essential!

Conclusions

- Exemplar models like DOP-LAST have words and structures stored in the same way
- Therefore lexical access and structural access should be the same
- Structural priming behaves like lexical priming:
  - Prime frequency
  - Prime-target similarity
  - Prime neighborhood density

Conclusions

- First model that predicts structural priming would be affected by similarity and neighborhood density
- First psycholinguistic evidence for exemplar models in syntax

Future work

- More modeling to make sure all three effects are independent and simultaneous
- Experimental confirmation
- Implement DOP-LAST

Thanks

- Florian Jaeger
- Joan Bresnan
- Tom Wasow
- Vsevolod Kapatsinski
- Rens Bod and Dave Cochran
- **And *you*!**