Evidence for a Unified Theory of Structural and Lexical Priming

Neal Snider

CUNY 2008

March 14, 2008

Lexical vs. structural priming

- Separate literatures
  - Lexical (Scarborough et al 1977; Versace and Nevers 2003; Ratcliff and McKoon 1981; 1988; Schreuder et al 1983; Perea and Rosa 2000; Anaki and Henik 2003; etc.)
  - Structural (Bock 1986; Bock and Griffin 2000; Pickering and Branigan 1998; Scheepers 2003; Hartsuiker et al 2007; etc.)
- Increasing evidence of storage of super-lexical "chunks" (Bod 2001; Schmitt and Galpin 2004; Schmitt, Grandage, and Adolphs 2004; Tremblay et al 2007)
  - may unify lexical and syntactic processing (MacDonald et al 1994; Jurafsky 1996)
- Might structural priming have same mechanism as lexical priming?

Outline

- 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
- Conclusion

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

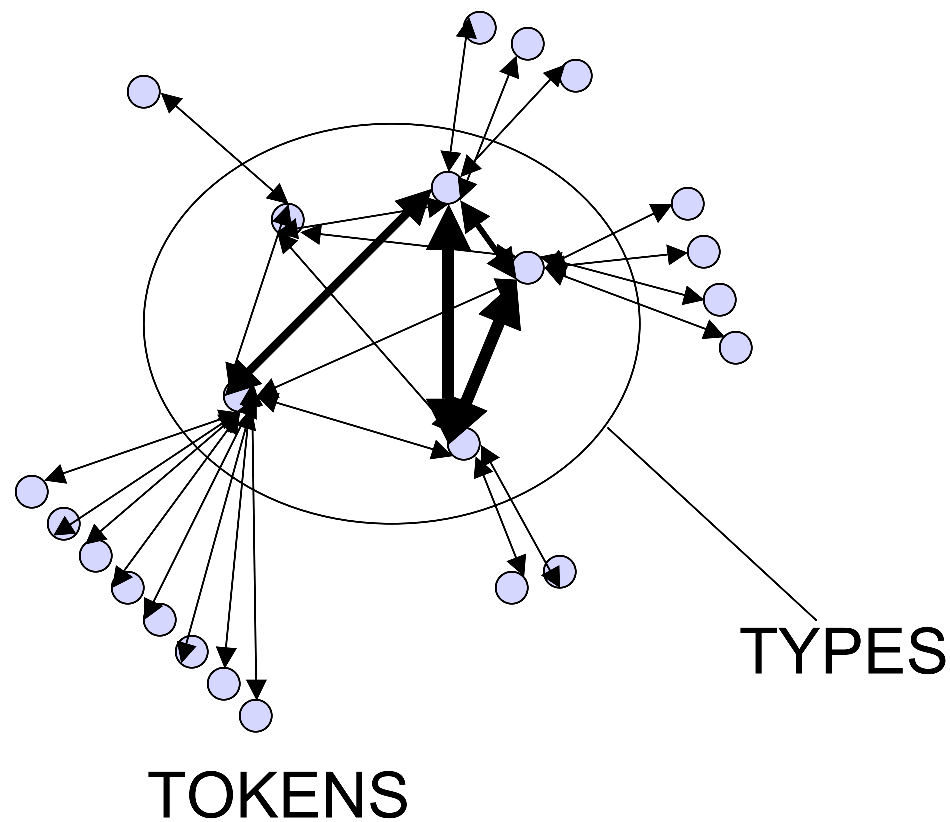
- Lexical access modeled as the spread of activation through a network of nodes
  - Nodes correspond to linguistic entities (words, phones, etc.)

- Used with success in phonology and morphology (McClelland and Rumelhart 1981, Anderson 1983)

LAST model (Kapatsinski 2006)

- Local Activation Spread Theory

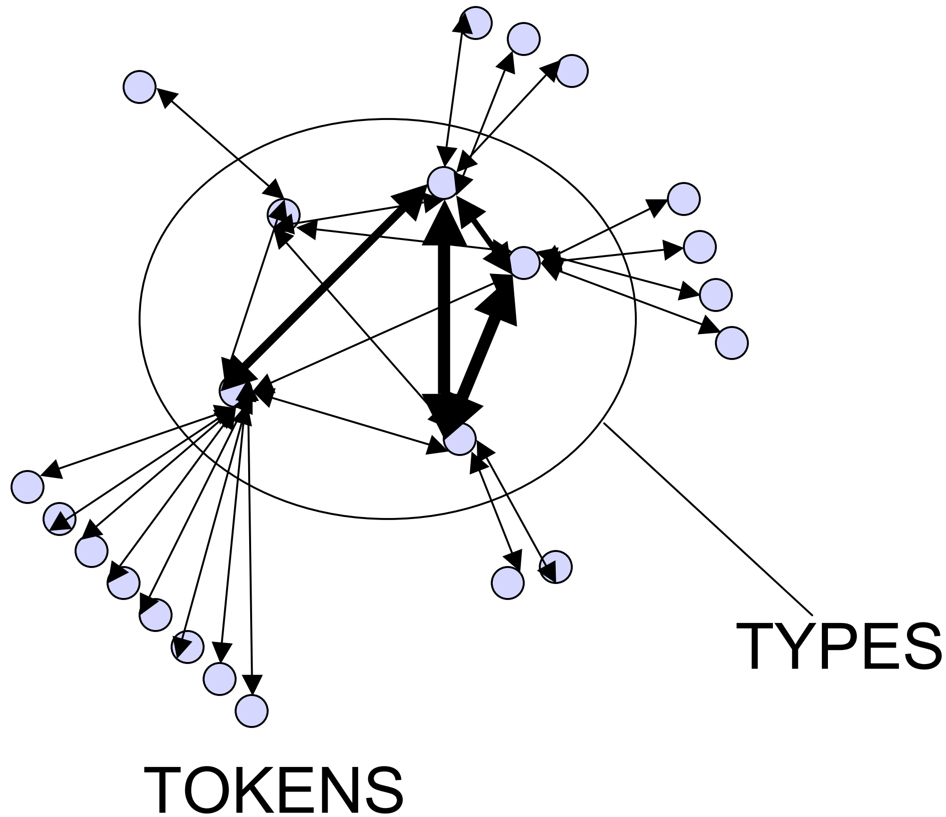
- Memory is a network
- In this network, each unit corresponds to a node



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

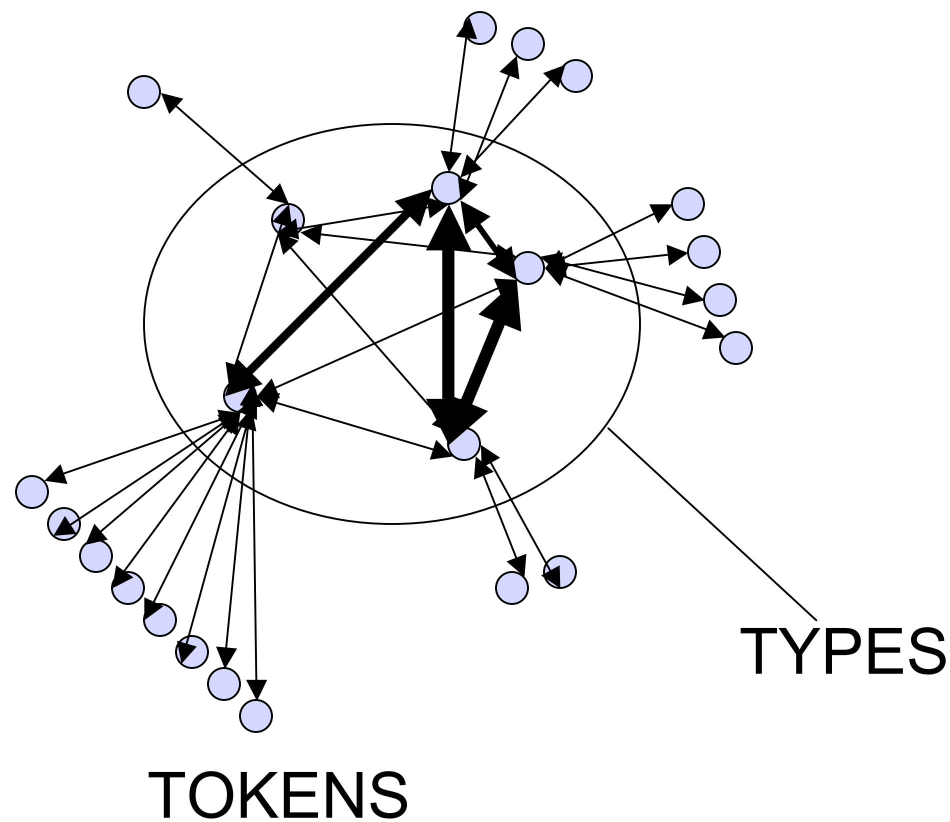
- Similar to HPSG/Construction Grammar type hierarchy



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



Syntactic LAST model

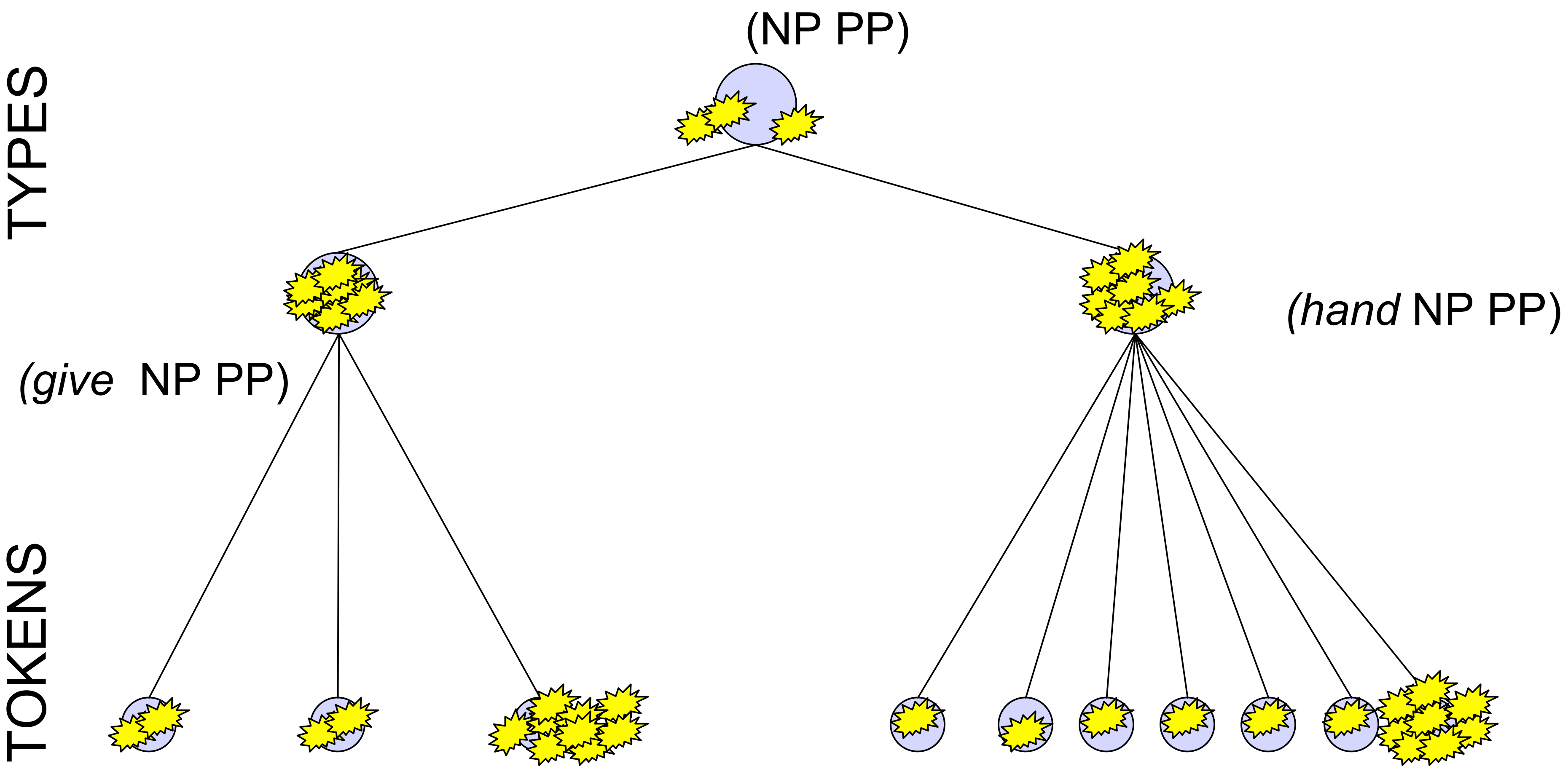
- Construction types can be acquired using exemplar methods like Data-Oriented Parsing (Bod 1992; 2006)
- Types can be connected in a network based on co-occurrence
- What does this predict for structural priming?

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

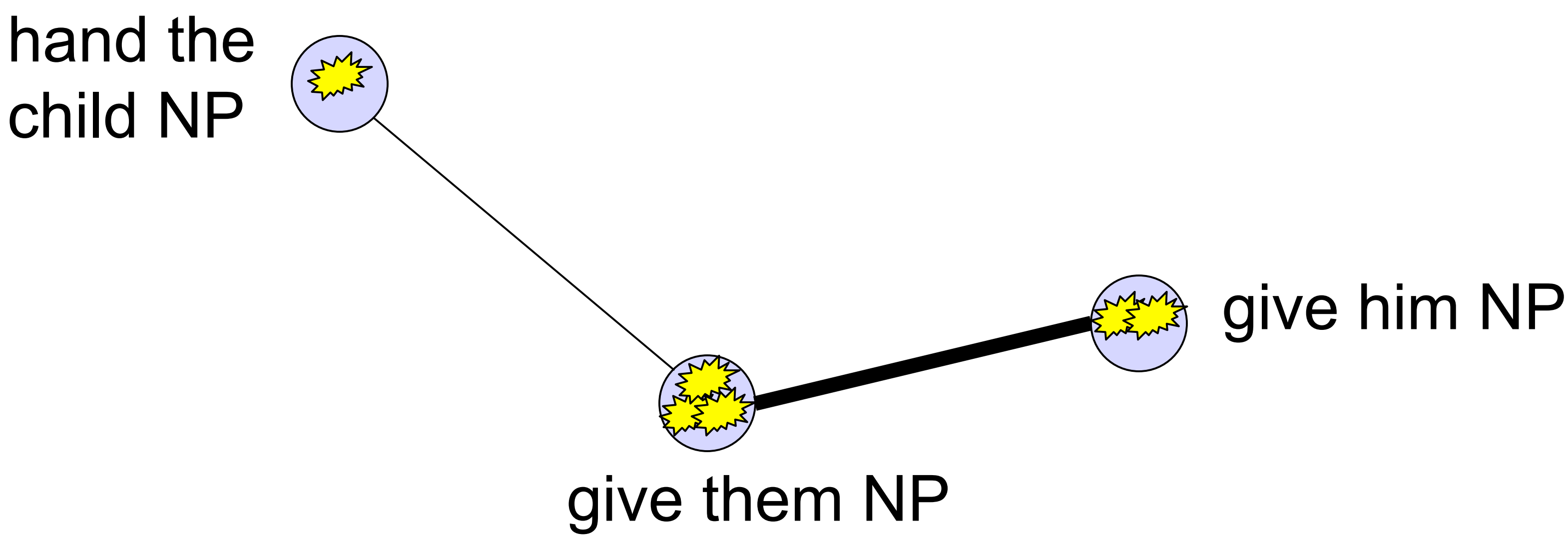
- How does LAST predict this?

Prime token frequency



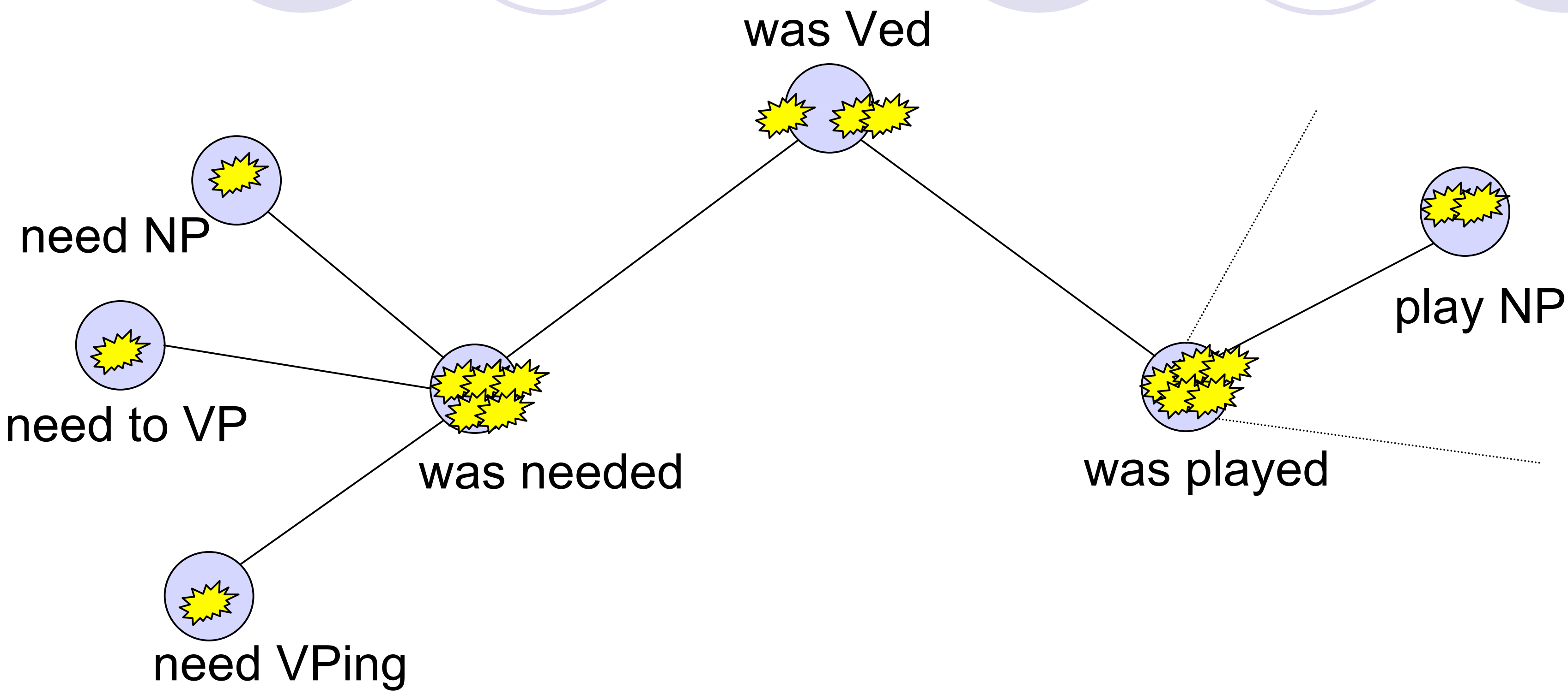
Prime types with more tokens leave less activation

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

- Words and constructions stored similarly
- Syntactic and lexical access can both be modeled by spreading activation networks
- Therefore, structural priming should behave like lexical priming

Data: Corpus study of priming in voice alternation

Problem

- Experimental studies of priming involve alternation of active and **full** passive:

  A fan punched the referee

  The referee was punched by a fan

  - Minimizes semantic priming
- Most passives (120/557) in corpus are reduced:

  The referee was punched
- How to make corpus data parallel experimental **syntactic** priming 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 sub-corpus (all primed turns)
- Impersonal actives acquired through information status coding (Nissim et al 2004)
- 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

Replication of Jaeger and Snider (2007)

Token frequency

- In this study, token 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

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
  - Found in corpus study of ditransitive (Snider and Jaeger 2007) and experimentally (Bernolet and Hartsuiker 2007)
- As predicted by spreading activation model
  - 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 structures should be more likely to prime!
- How to measure similarity?

- Nearest-neighbor models offer precise metric of 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 structures that have less 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
- Found in ditransitive priming (Snider 2007)
  - Also see Reitter (2008)

Summary so far

- Just as in lexical priming:
- Structural priming is affected by
  - (Inverse) frequency
  - Similarity

- Evidence that lexical production and structural production access same representations
  - As predicted by construction network 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 network model
- Demonstrates necessity of fully-connected **types**
  - abstraction is essential!

Conclusions

- Network models like LAST process words and structures 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

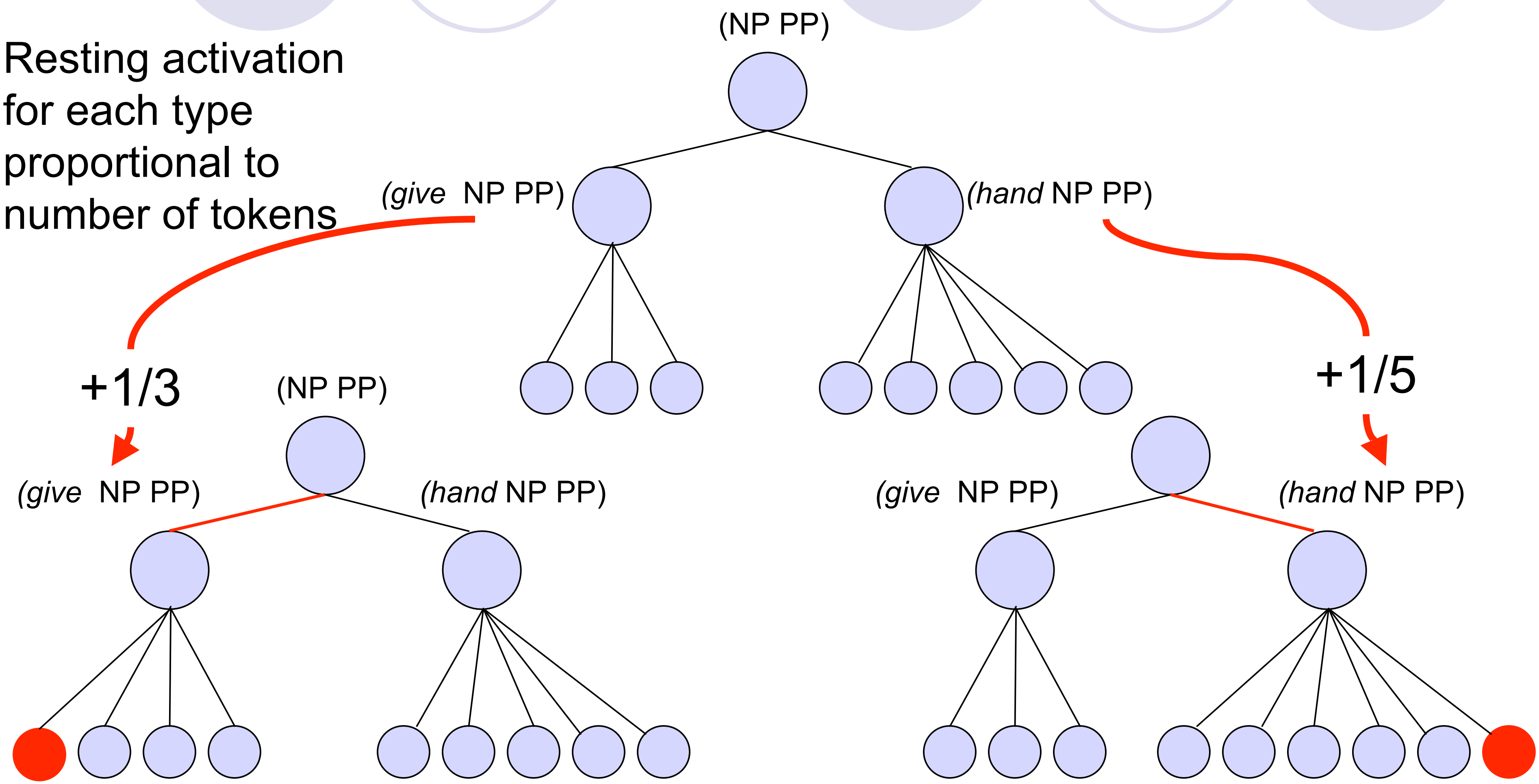
Priming as implicit learning

- But what about evidence that (some) priming is long lived and involves learning? (Bock and Griffin 2000; Chang et al 2003; Jaeger and Snider 2007; Reitter et al 2006)

- Network models like LAST that involve token storage predict another route for priming effects
  - So does ACT-R model of Reitter and Keller (2008)

Another route: learning

Resting activation for each type proportional to number of tokens



Less frequent type has more increase in resting activation after 1 token

Activation vs. learning

- Model predicts two routes for priming effects
- Different implications:
  - Learning: token frequency effects should be long-lived
  - Activation spread: similarity and neighborhood density effects should more sensitive to prime-target distance
    - Verb similarity boost is indeed short-lived!

(Hartsuiker et al 2007)

- More experiments needed

Conclusions

- Evidence that lexical and structural priming are be similar processes
  - Of course, still many differences!
- Further support for chunk storage
  - Exemplar (Bod 1992; 2006) and fine-grained construction models (Goldberg 2004)
- First model that predicts structural priming would be affected by neighborhood density

Future work

- More modeling to make sure all three effects are independent and simultaneous
- Experimental confirmation and testing of differential decay effects

- Implement LAST to tease apart learning and activation mechanisms

Thanks

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