Domain General and Domain Specific Mechanisms in Real-Time Grammatical Computation

Colin Phillips

Linguistics Department
Language Science Center
Neuroscience & Cognitive Science Program

UNIVERSITY OF 18 56 MARYLAND

Language at Maryland














Matt Wagers (UCSC)
Ellen Lau (UMD)
Brian Dillon (UMass)
Nina Kazanina (Bristol)
Masaya Yoshida (NWU)
Shota Momma (UMD)
Dave Kush (Haskins/Yale)

Sol Lago (Potsdam)
Akira Omaki (JHU)
Shevaun Lewis (JHU)
Wing Yee Chow (UCL)
Jon Sprouse (UConn)
Alexis Wellwood (NWU)
Dan Parker (Wm & Mary)

“The function of a symbol is is to carry information forward in time in a manner that makes it accessible to computation.”

(Randy Gallistel, yesterday morning)

“My goal is to understand computation using structured representations in memory.”

(Colin Phillips, very early yesterday morning)

Fast

Language Processing

Process models
Focus on what speakers do quickly

Little interest in abilities that manifest more slowly

Grammatical Theories

Process-neutral models
Focus on what speakers can do
when freed of time/memory limits

Talk of 'mental computation'
But little interest in real-time operations

Slow

Not a linguistics vs. psychology split:

Language development shows no such divide

Map

0. Levels of analysis

1. Reconciling linguistic constraints & memory mechanisms

2. Reductionism vs. implementation

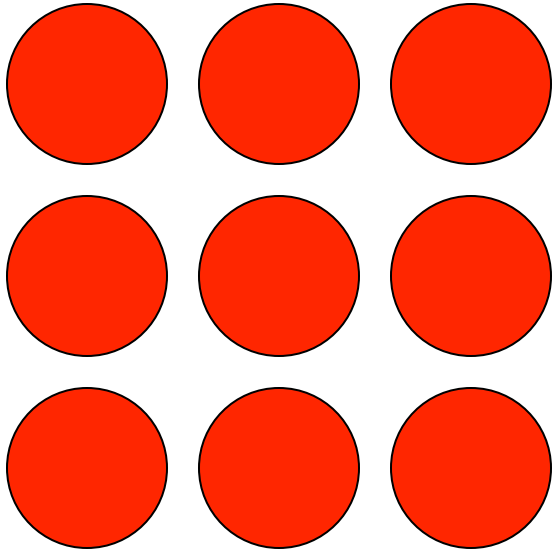
3. Linking linguistic cues to world knowledge

Levels of Analysis

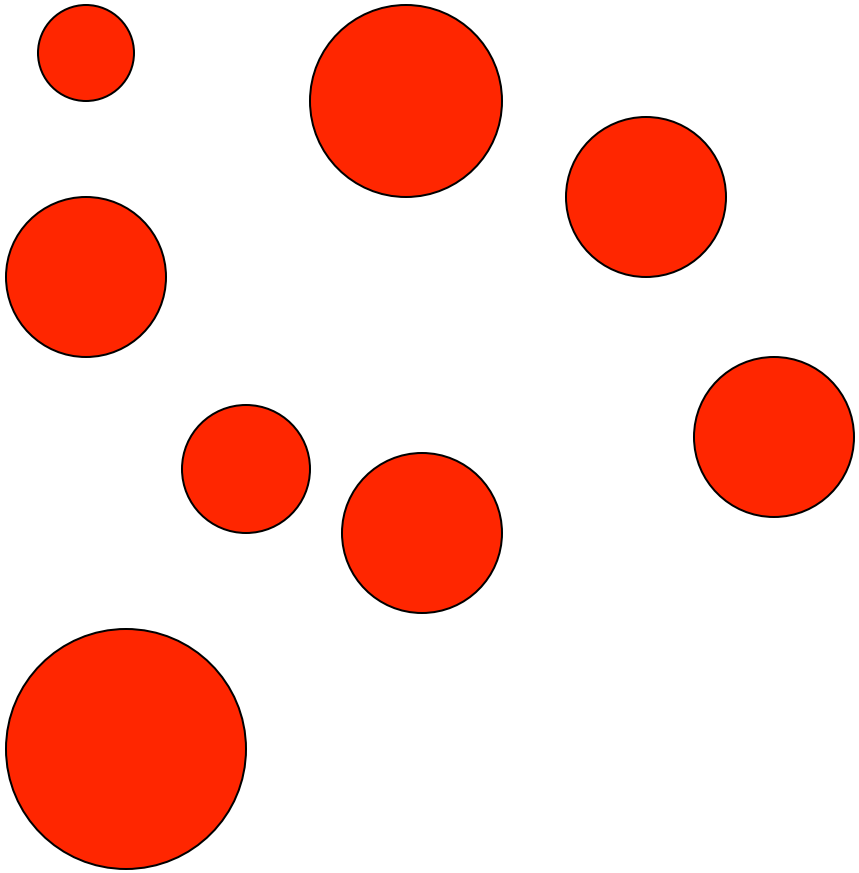
≠

Architectural Components

how many dots?



how many dots?



Number Cognition

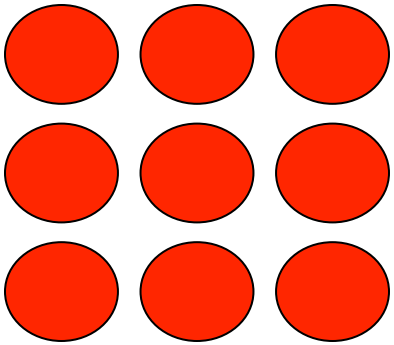
- Multiple systems for encoding quantity:

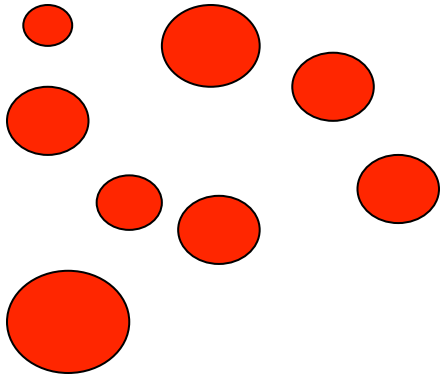
  Approximate number system
  Exact number system

- **Task-specific routines** for specific numerical problems can be distinguished from **deeper understanding** of number

  e.g., 423 x 56 = ___

  e.g., early counting: “one, two, three, four … two dogs!”





Chomsky 1965

- “We thus make a fundamental distinction between competence (the speaker-hearer's knowledge of his language) and performance (the actual use o r the idealization mance a direct reflect Observed us to the nature of this mental reality, but surely cannot constitute the actual subject matter of linguistics, if this is to be a serious discipline.” (p. 4)

claim #1
cognitive system vs. behavior
(necessary)

- “To avoid what has been a continuing misunderstanding, […] a generative gra earer. […] When we say with respect to a particular ut how the speaker or he fficient way, to construct su to the theory of language use - the theory of performance. (p. 9)

claim #2
different cognitive systems
(empirical hypothesis)

Standard Grammatical Analysis

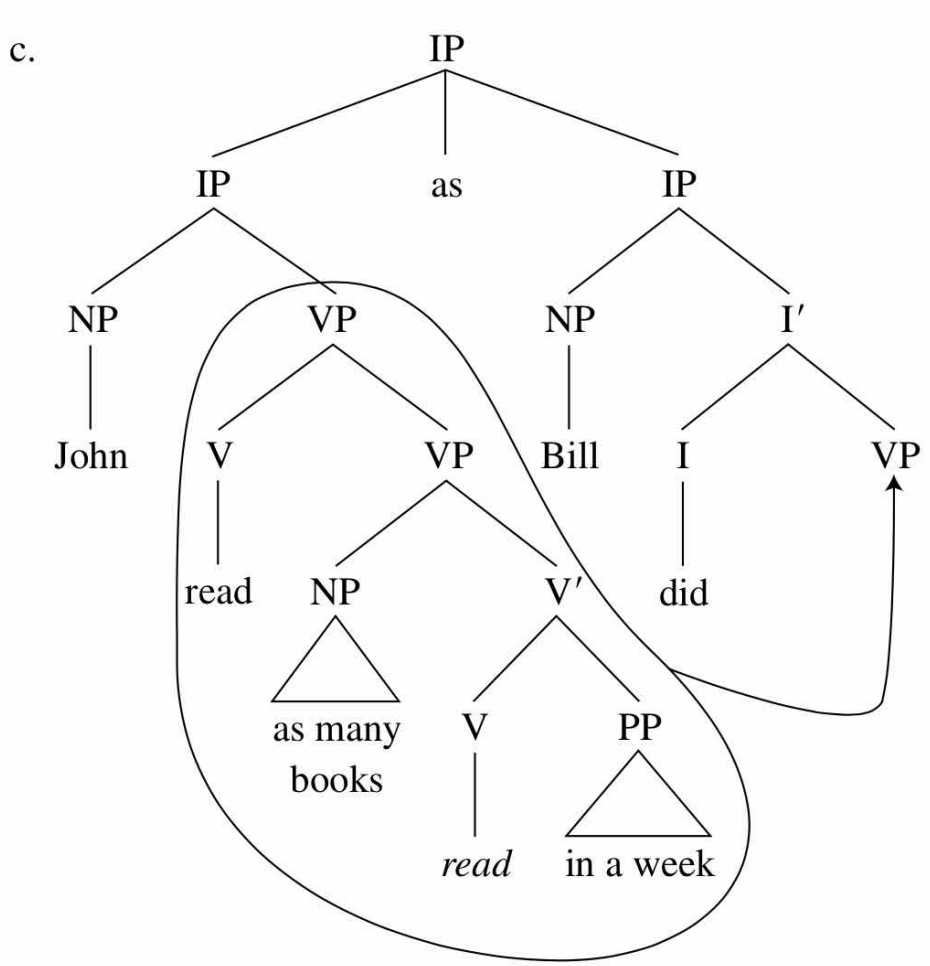
(a.k.a. 'syntactic theory')

Hierarchical groupings of terminals

All elements are discrete symbolic representations

No time dimension

Derivations generally not taken as claims about actual time steps
(Phillips & Lewis 2013)



Default questions:
How acceptable is this sentence?
Why is it so (un)acceptable? Does it violate combinatorial rules? Is it just hard?

… This is a narrow set of questions to ask.

Cognitive Models of Sentence Structure Building

(a.k.a. 'processing theories')

Same questions + many more

Order always matters
Time sometimes matters

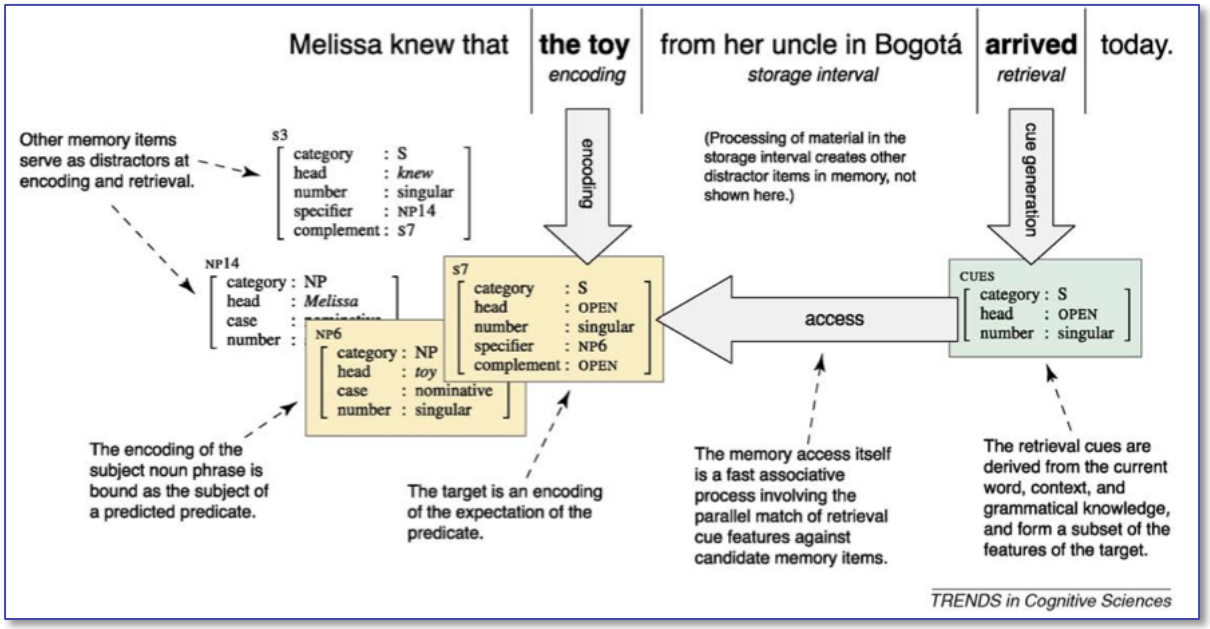
representation =
memory encoding

dependencies =
memory access

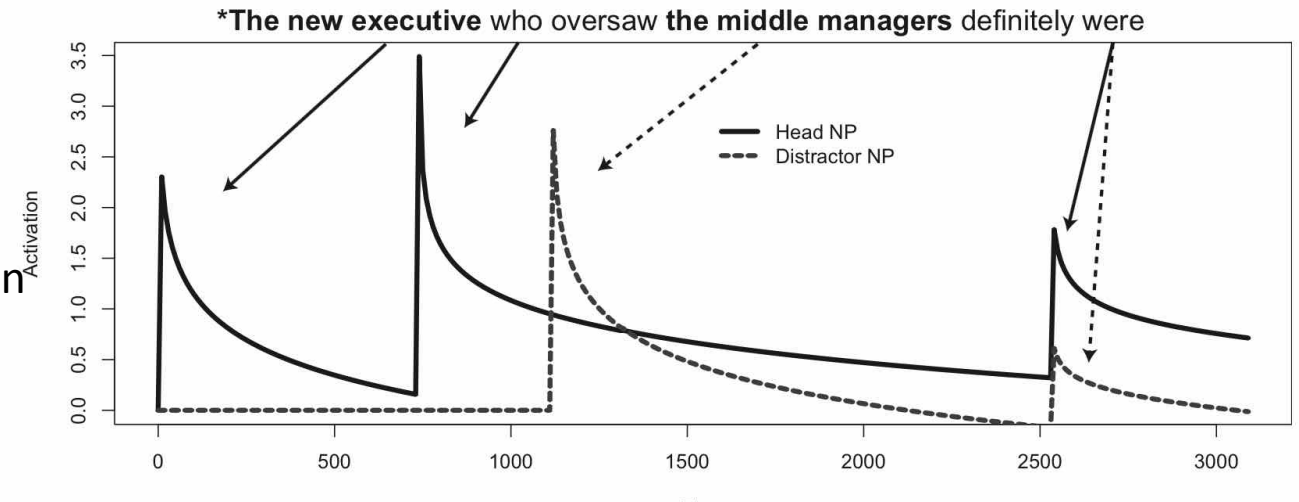
nodes may have varying activation levels

computations may depend on indep. cogn. abilities

comprehension, production work w/ limited information



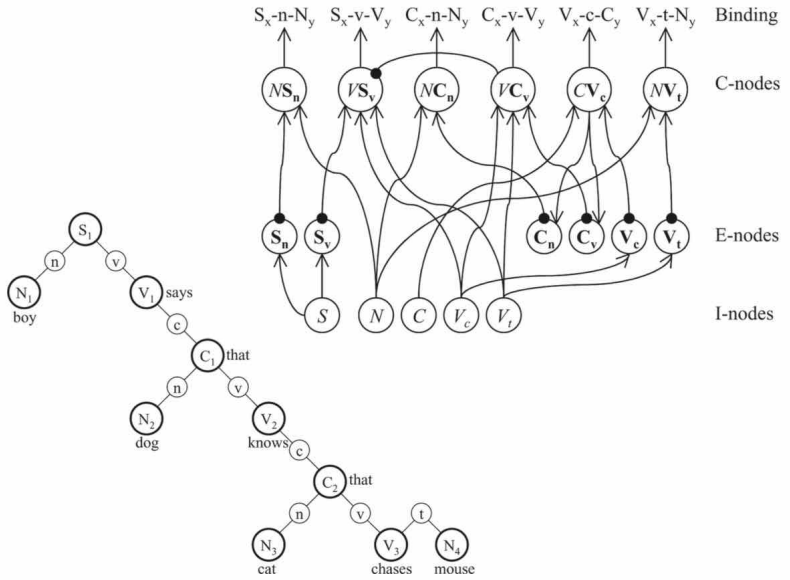
Lewis, Vasishth & Van Dyke 2006

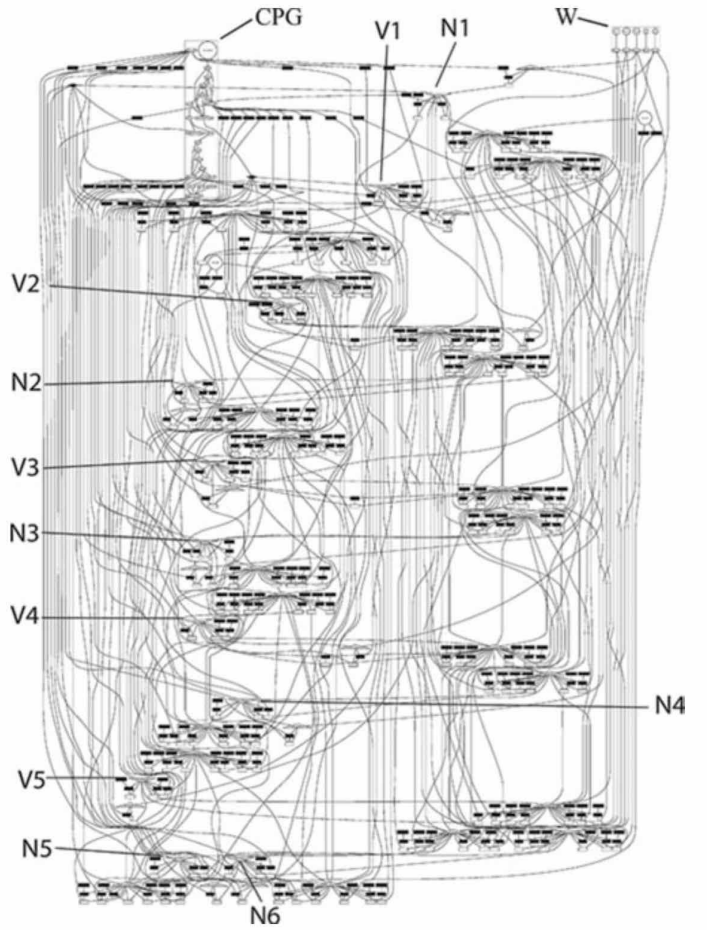


Dillon, Sloggett, Mishler, & Phillips 2013

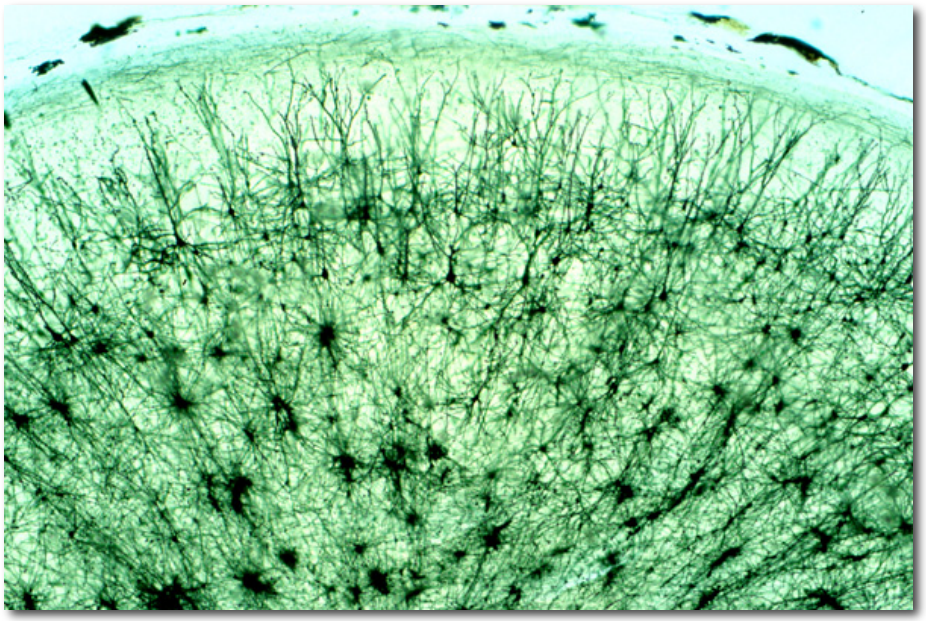
Neural Models of Sentence Structure Building

Different vocabulary / toolkit
Constraints from connectivity & anatomy
Nodes correspond to complex activity pattern

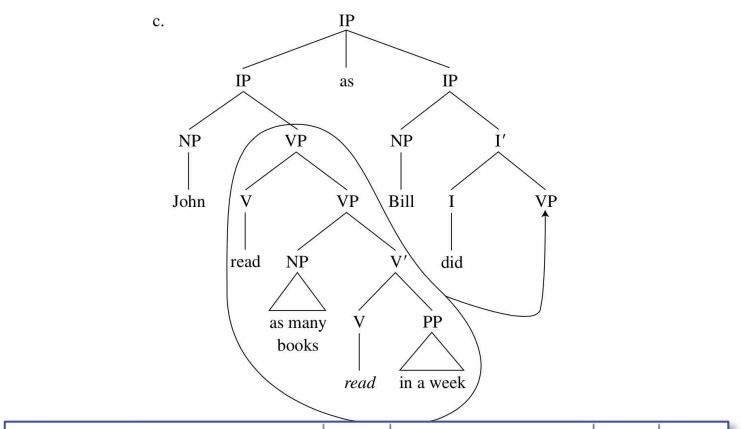




Van der Velde & de Kamps 2006

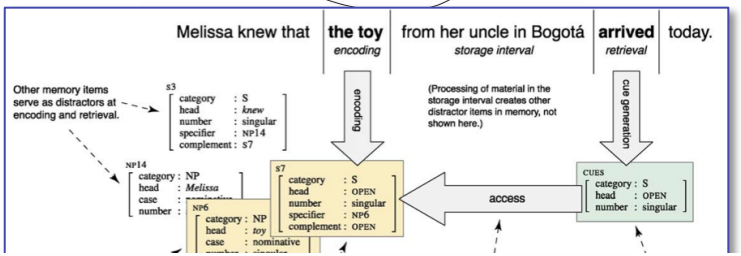


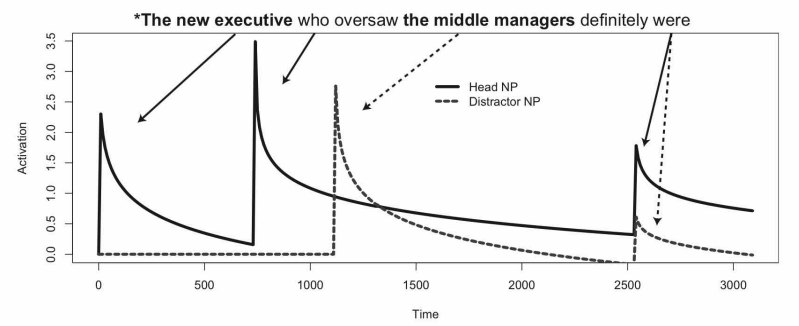
One Language System
Multiple Levels of Analysis



"linguistic"

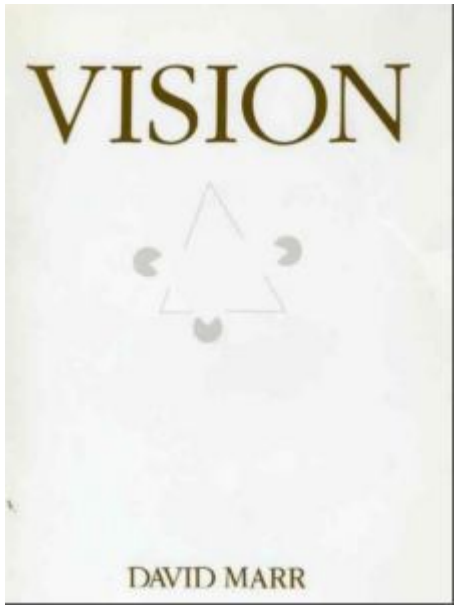
"computational?"



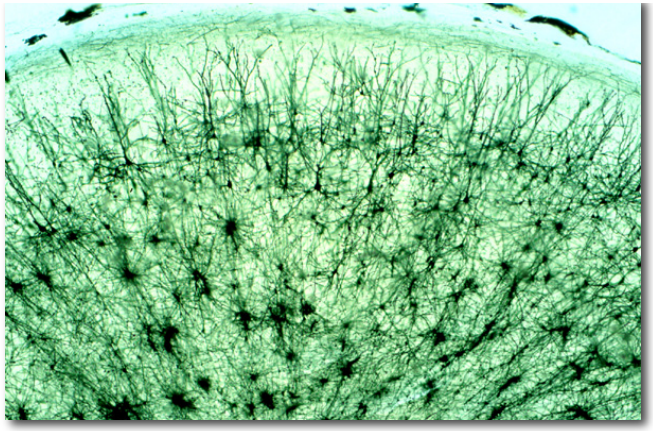


"cognitive"

"algorithmic?"



Marr 1982



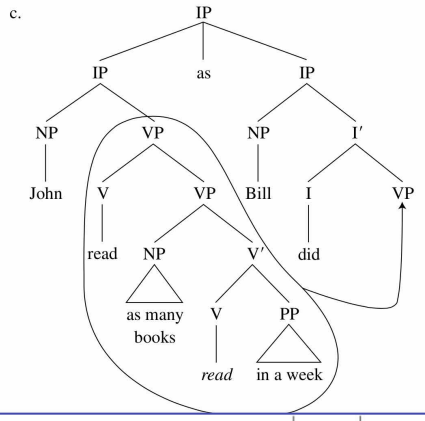
"neural"

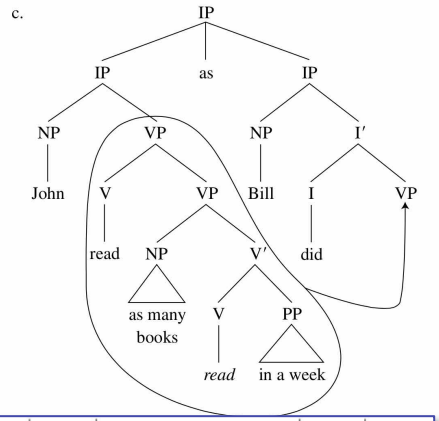
"implementational"

Marr-levels don't really fit language research.

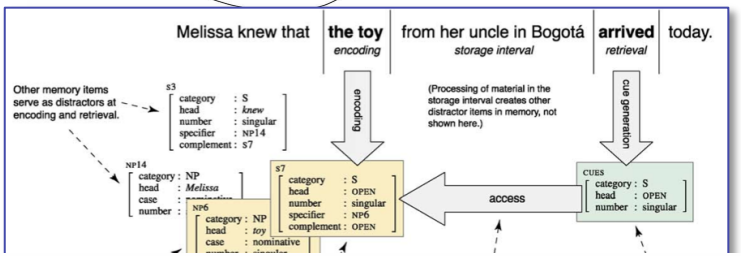
And there aren't really 3 discrete levels.

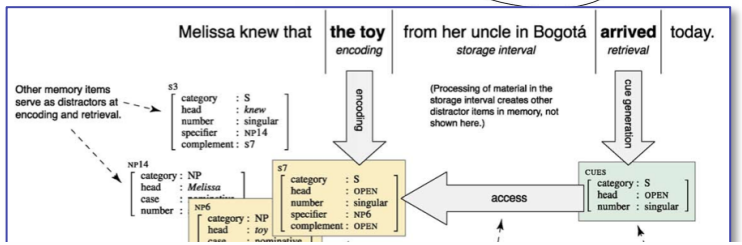
Multiple Language Systems
Multiple Levels of Analysis

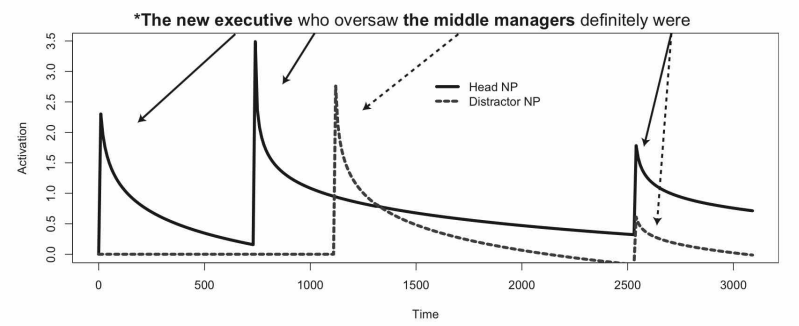


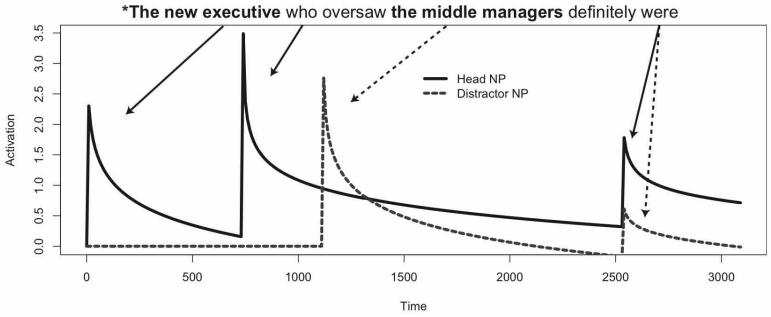


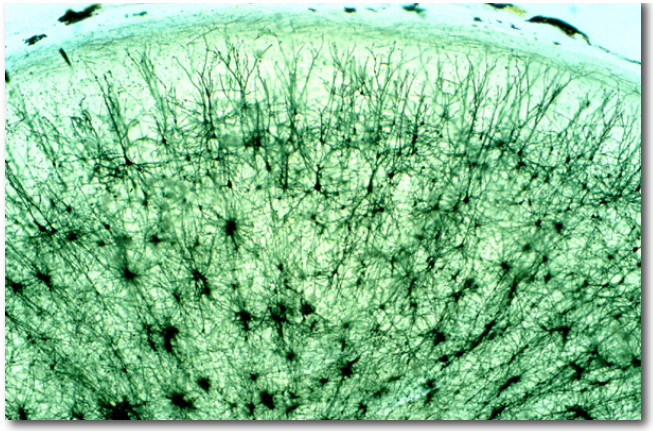
← Grammar
Parser →
Producer

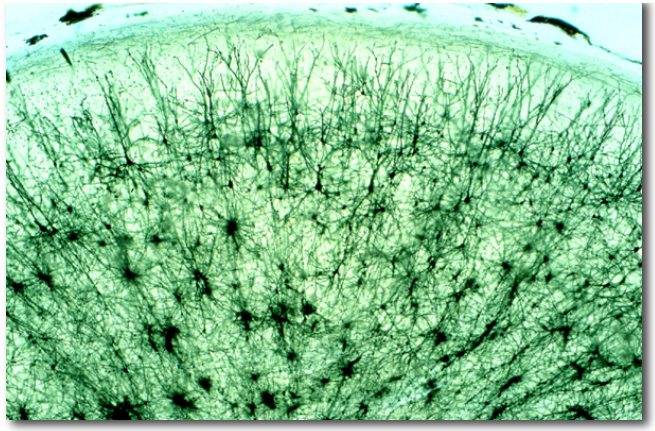


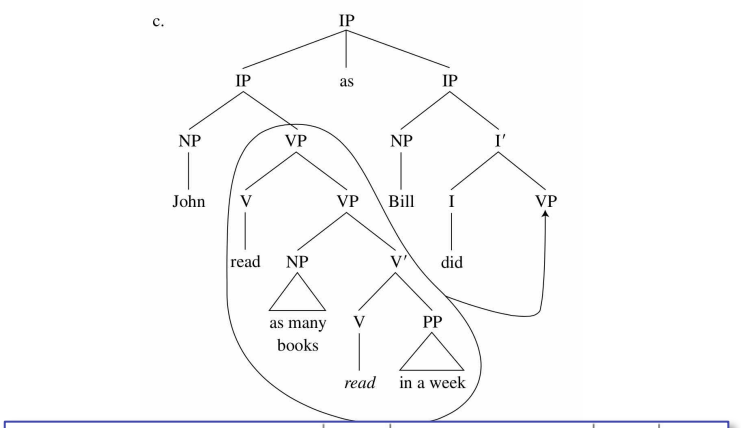






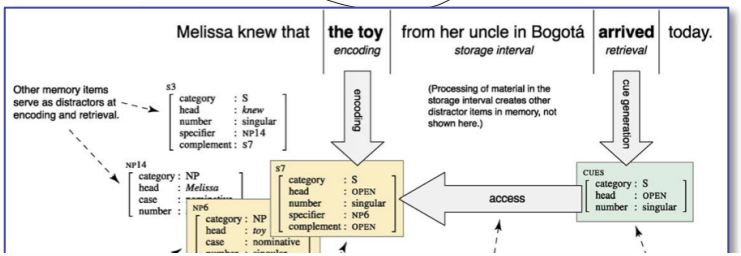






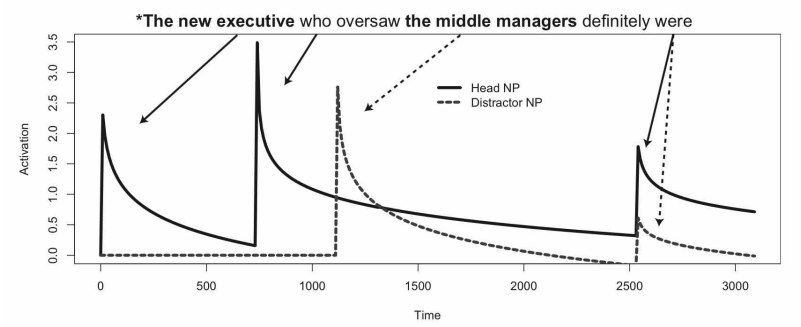
structural constraints

(left-right) order



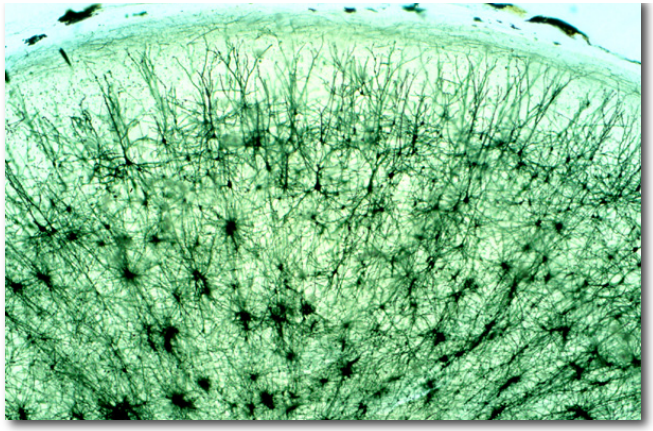
time

Blame Assignment



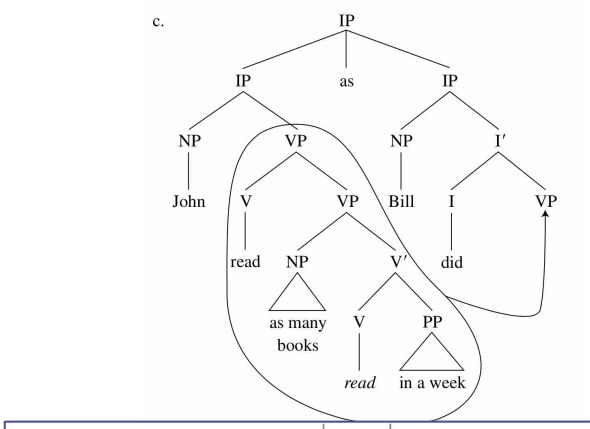
resource limitations

memory architecture and access



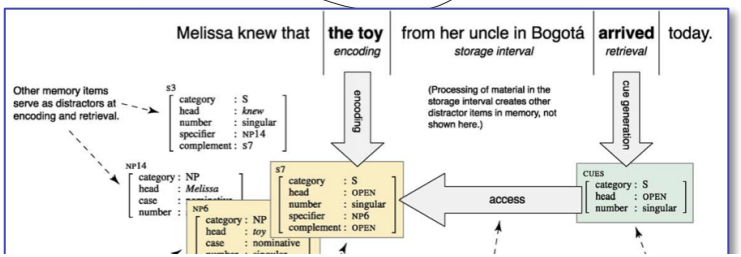
non-discrete units

etc.



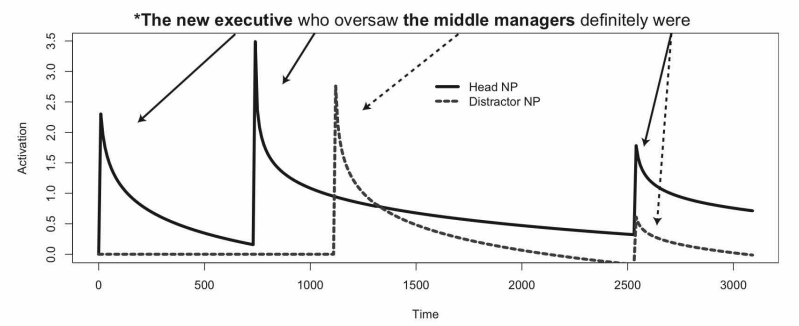
structural constraints

(left-right) order

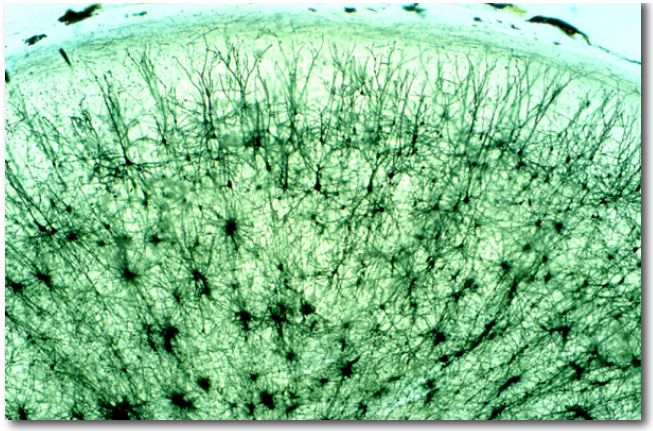


time

Standard Syntactic Accounts



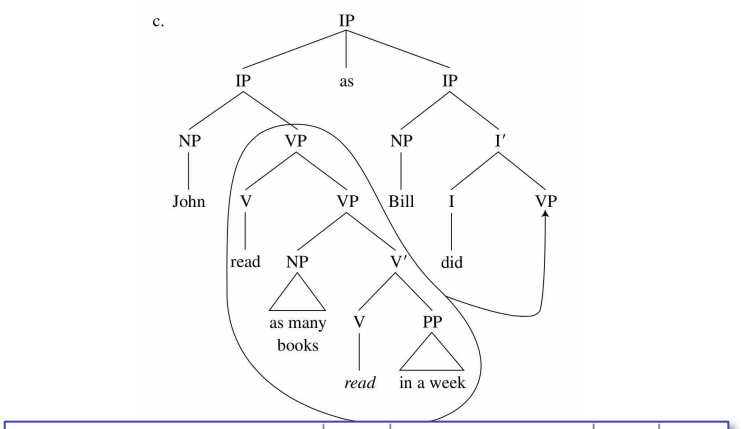
resource limitations



memory architecture and access

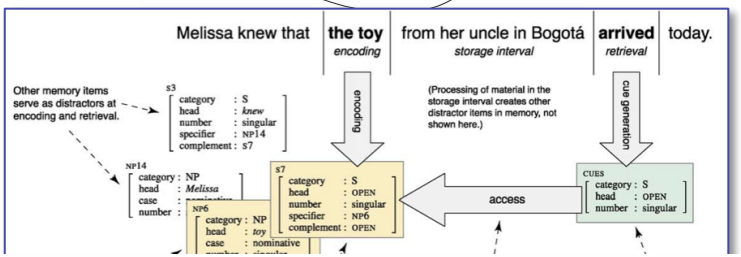
non-discrete units

etc.



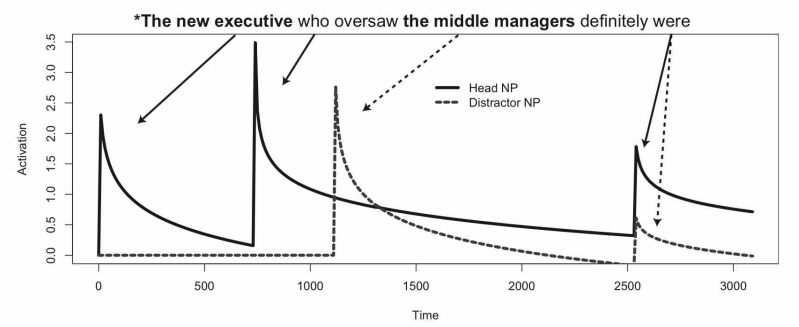
structural constraints

(left-right) order



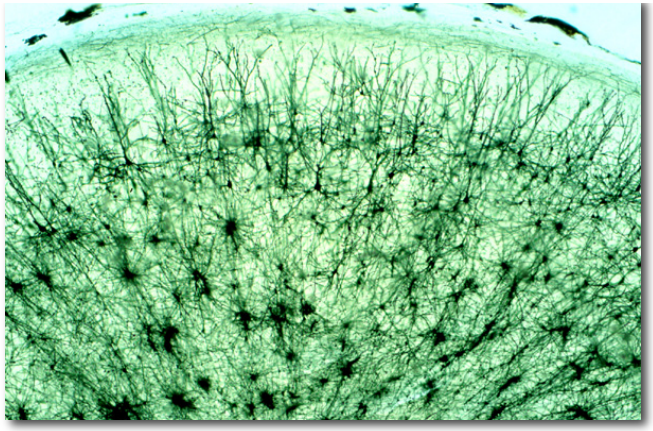
time

Reductionist Accounts



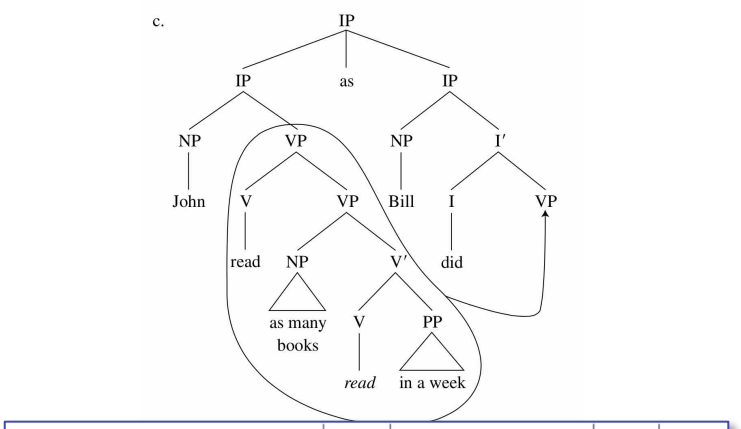
resource limitations

memory architecture and access



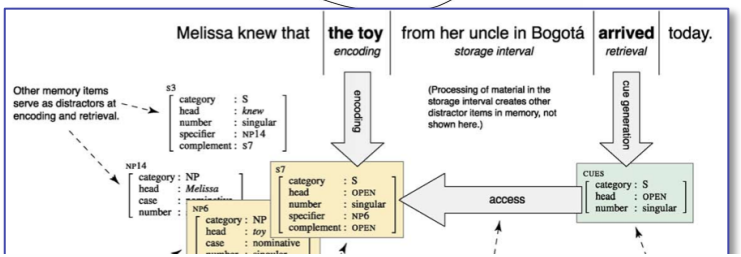
non-discrete units

etc.



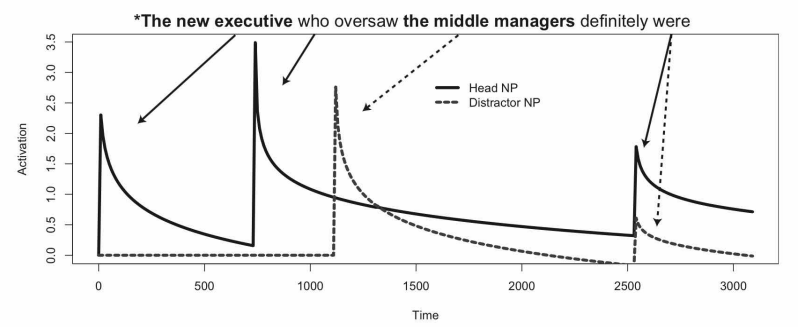
structural constraints

(left-right) order



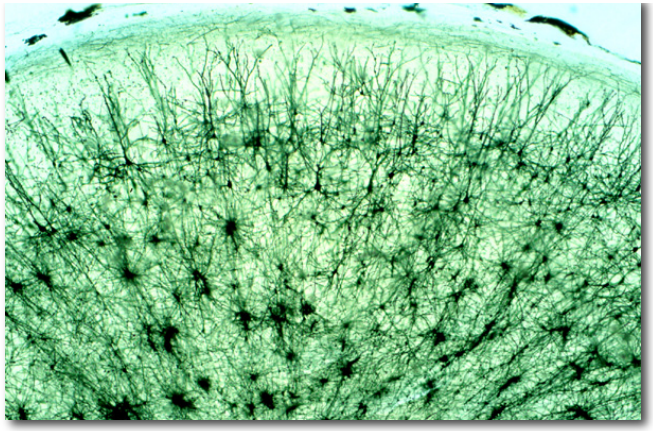
time

Enriched Accounts



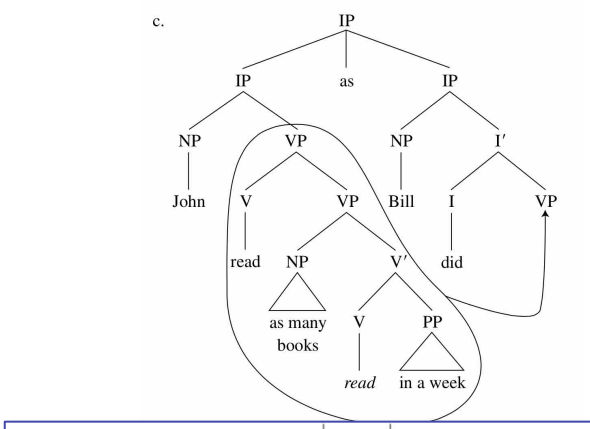
resource limitations

memory architecture and access



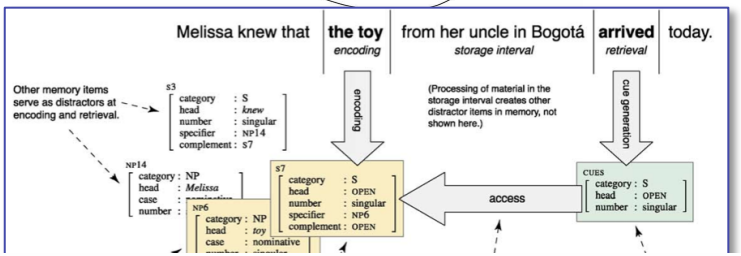
non-discrete units

etc.



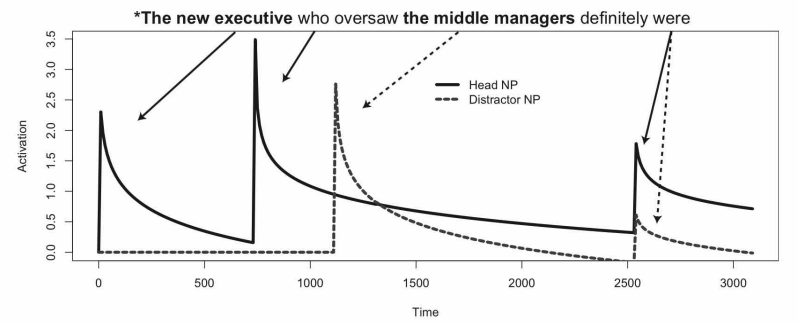
structural constraints

(left-right) order



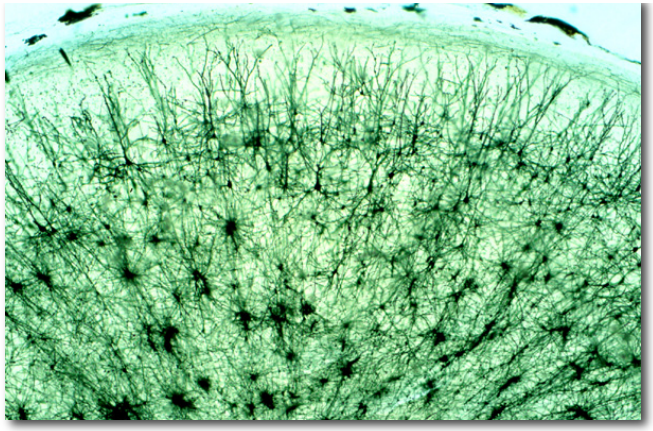
time

Order-based Accounts



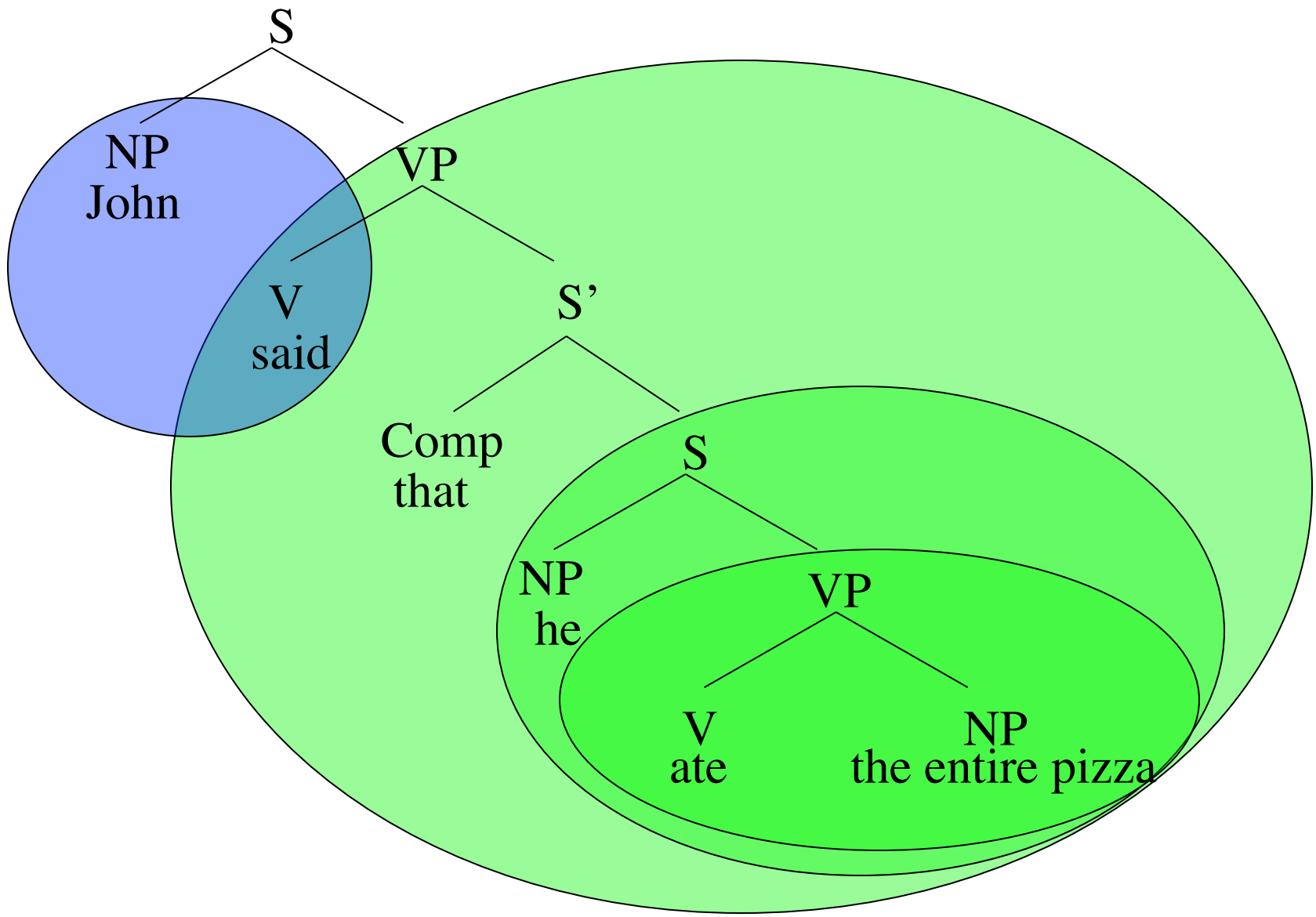
resource limitations

memory architecture and access

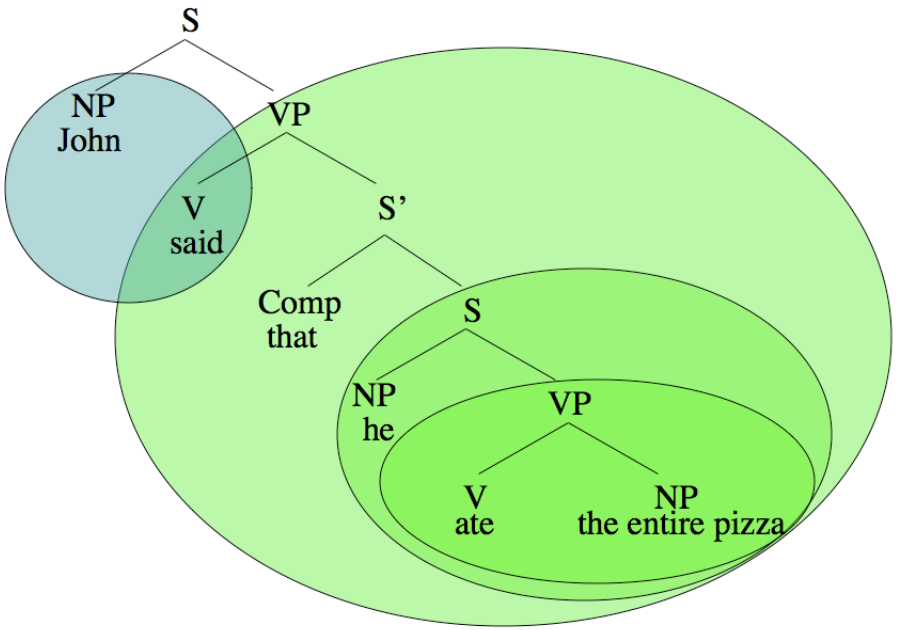


non-discrete units

etc.



An order-based account: Phillips 1996, 2003



- Many tools used to diagnose groupings of words:
  - coordination
  - deletion
  - interpretation (coreference, scope)
  - movement, focus, topicalization
  - etc.

- There are many cases where the tools converge on the same result
- There are also many cases where the tools yield conflicting results

- Suggestion: this well-known dilemma reflects **opacity effects** arising from left-right derivations

  (cf. classic arguments for derivations in phonology)

An order-based account: Phillips 1996, 2003

- Related arguments

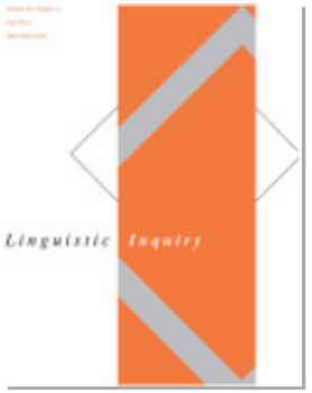
  - Schlenker (1999): German adjective inflection
  - Boeckx (1999): islands
  - Richards (2000): 'subjacency tax' effects in Japanese
  - Guimaraes (2003): syntactic amalgams
  - McGinnis & Gerdts (2003): Kinyarwanda applicatives
  - Barker & Shan (2006): crossover phenomena
  - Chung et al. (2006): licensing 'sprouting'
  - Choi & Yoon (2006): Japanese/Korean argument clusters
  - Kempson, Cann et al. (2001/2005): varied phenomena [*Dynamic Syntax*]
  - Steedman (2000, etc.): various phenomena [*Comb. Categorial Grammar*]

Subject Islands, Reconstruction, and the Flow of the Computation

Valentina Bianchi, Cristiano Chesi

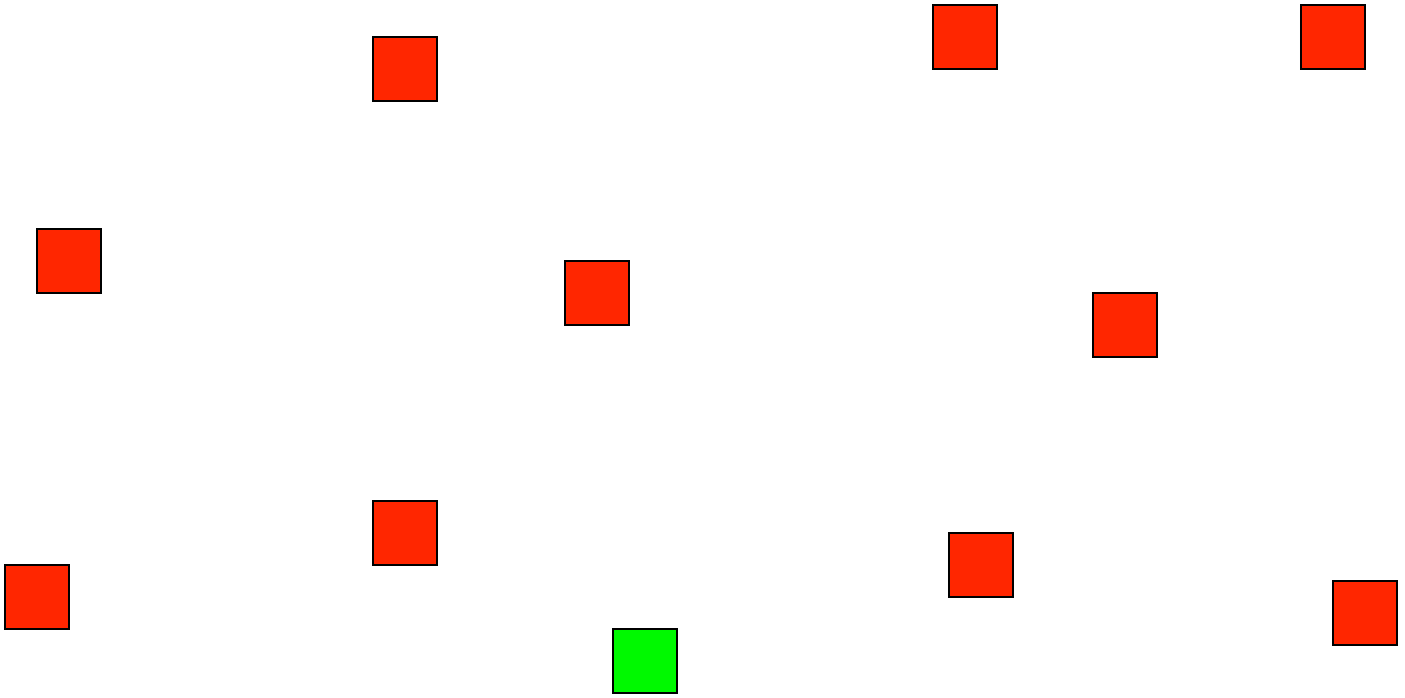
Linguistic Inquiry, Volume 45, Number 4, Fall 2014, pp. 525-569 (Article)

Published by The MIT Press

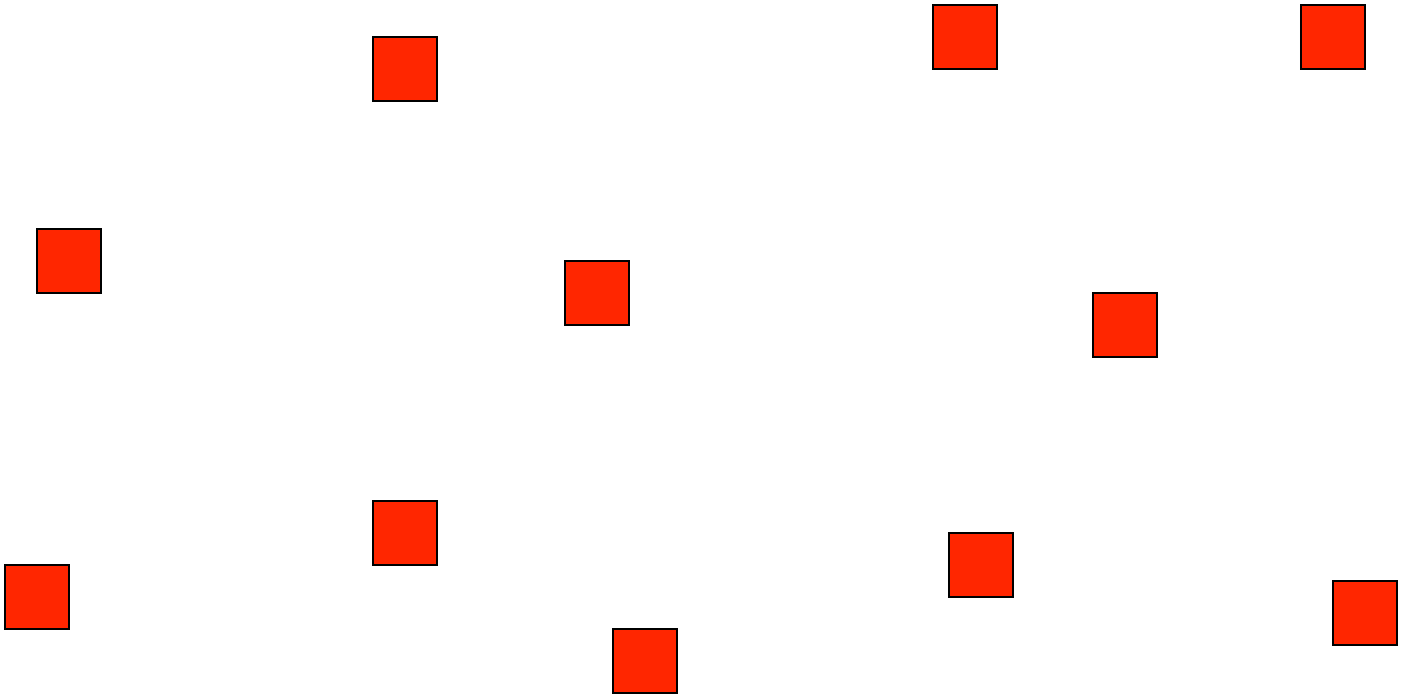


1. Reconciling Access Mechanisms & Linguistic Constraints

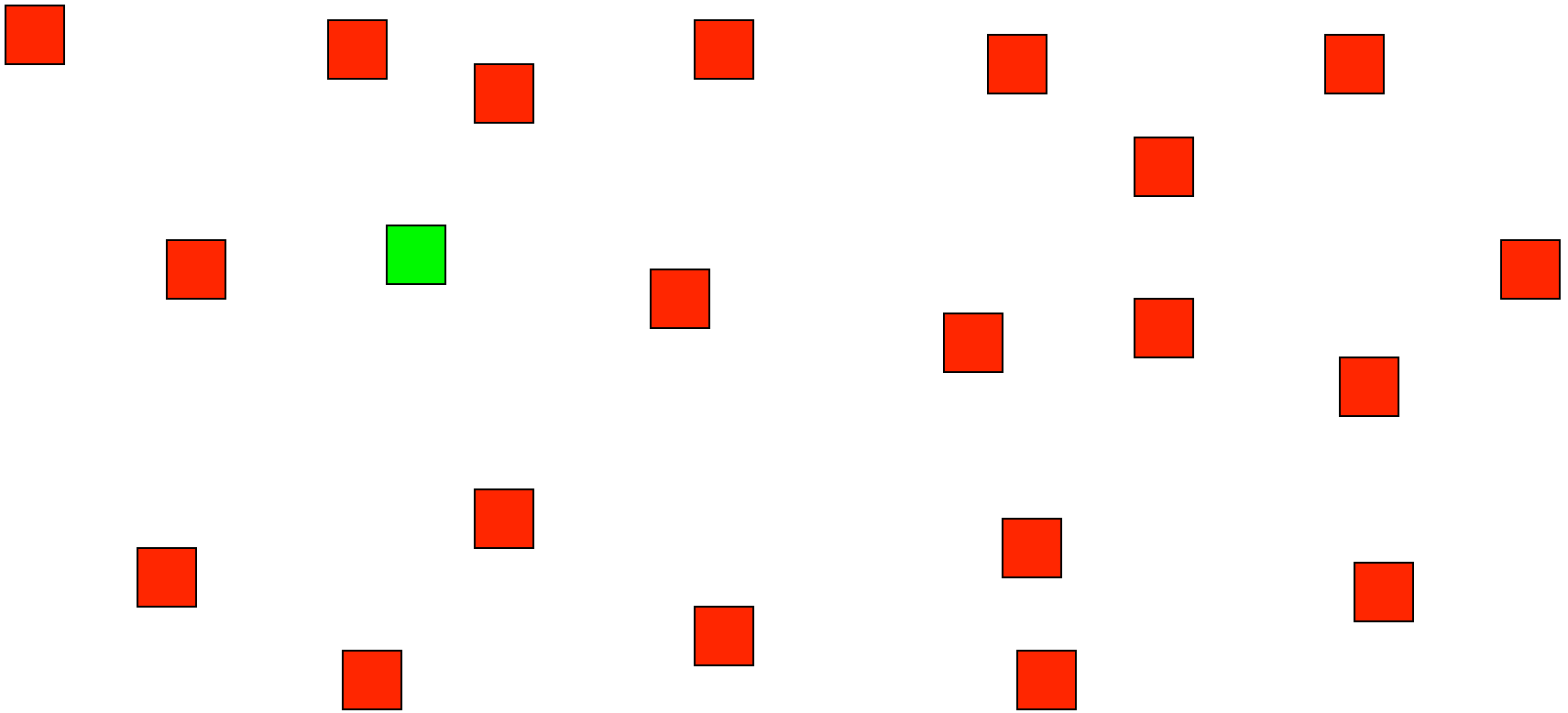
Is there a green square?



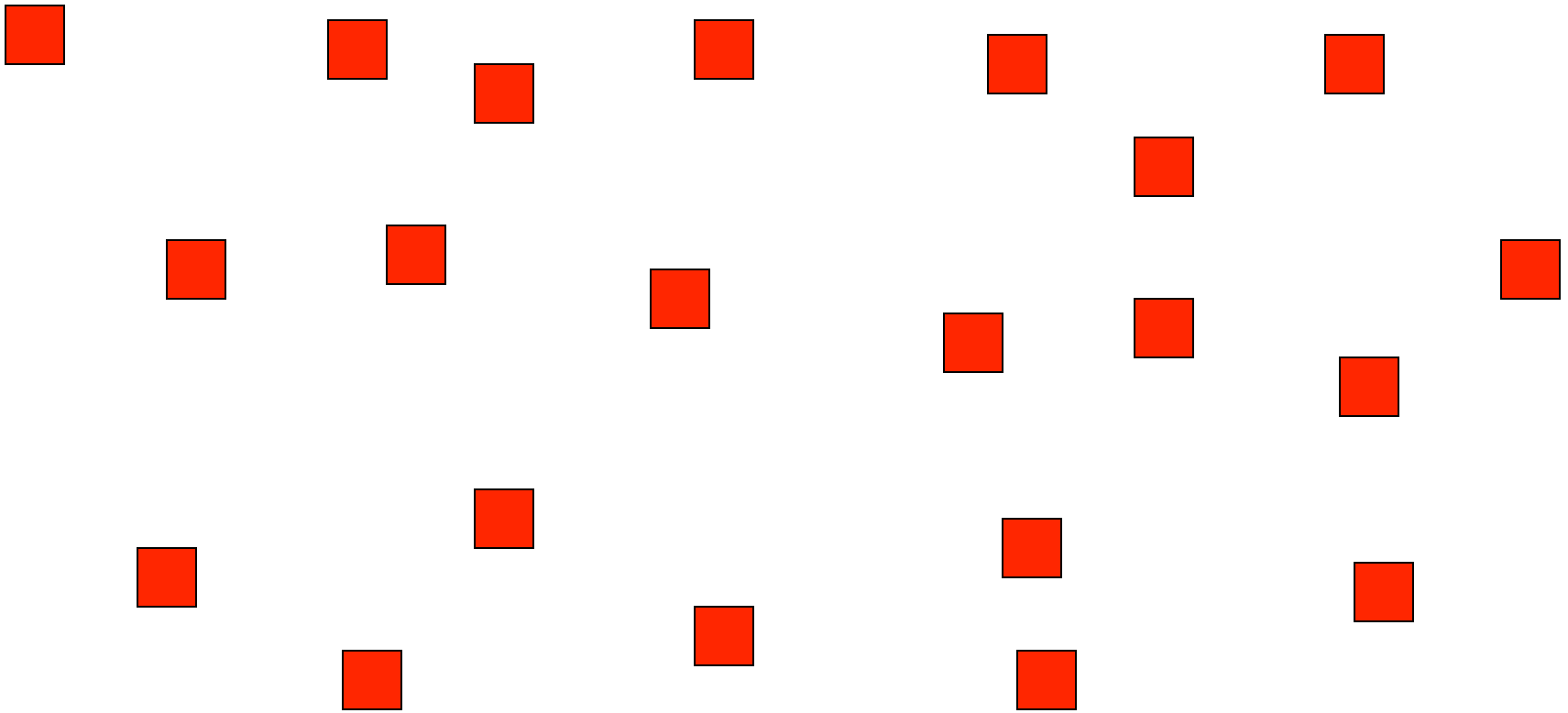
Is there a green square?



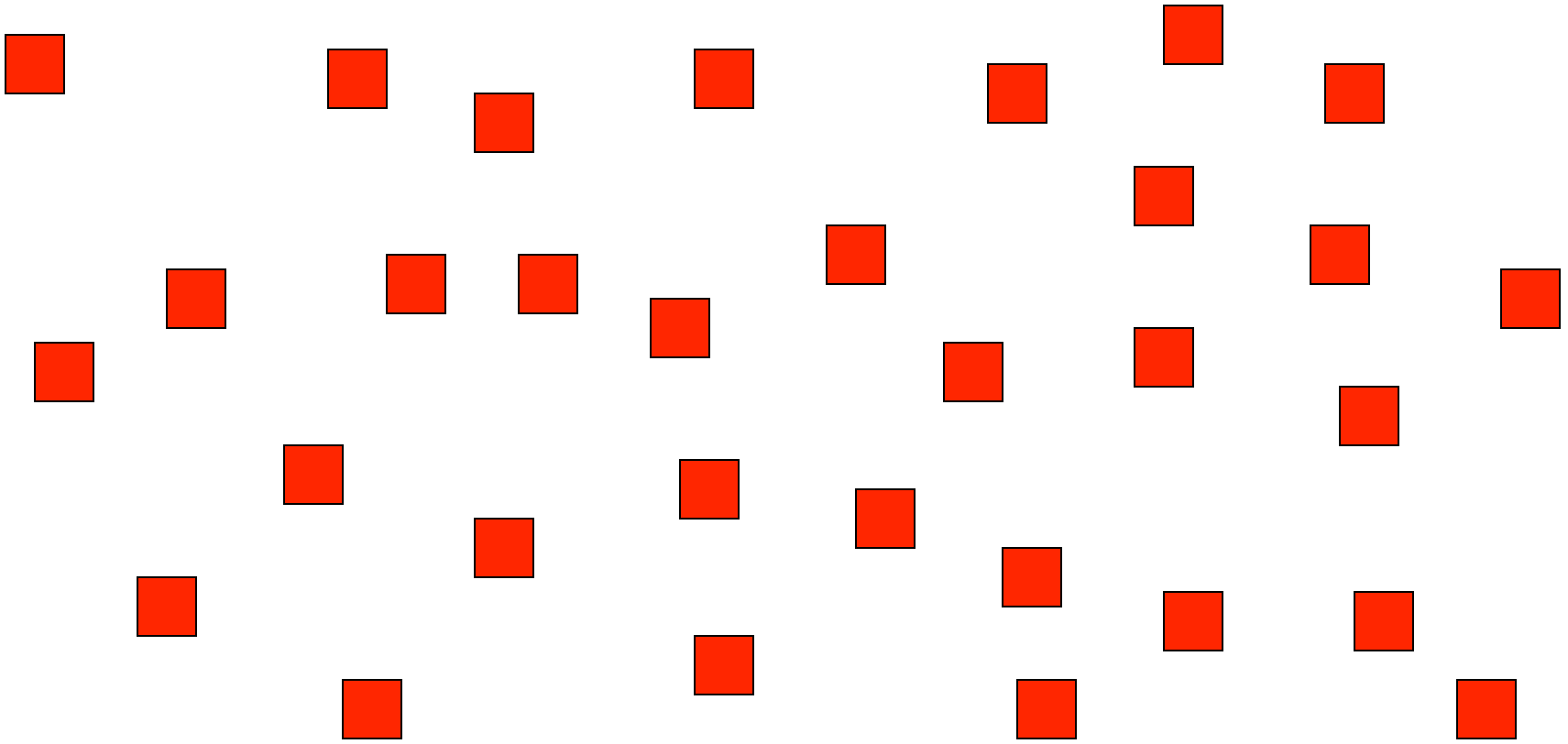
Is there a green square?



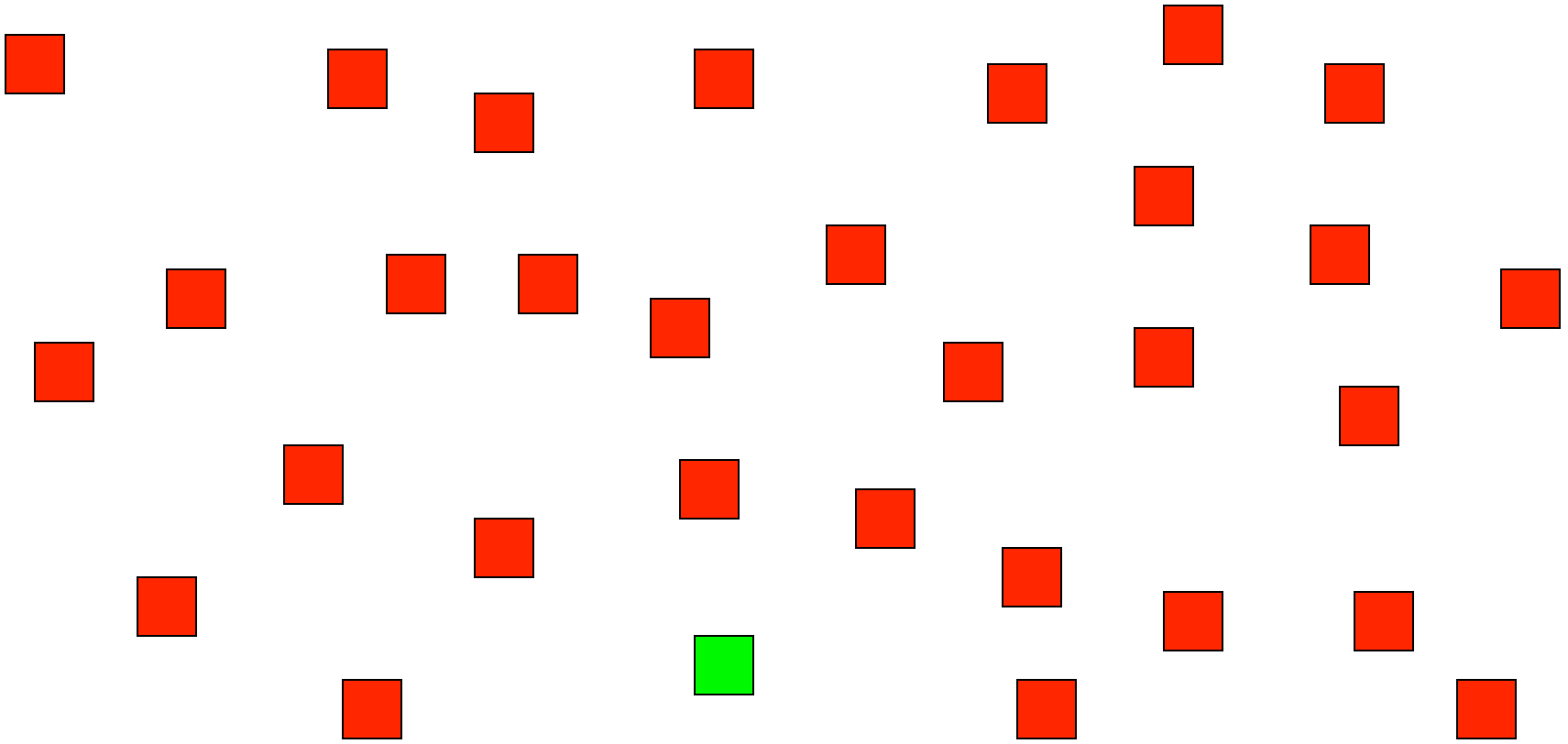
Is there a green square?



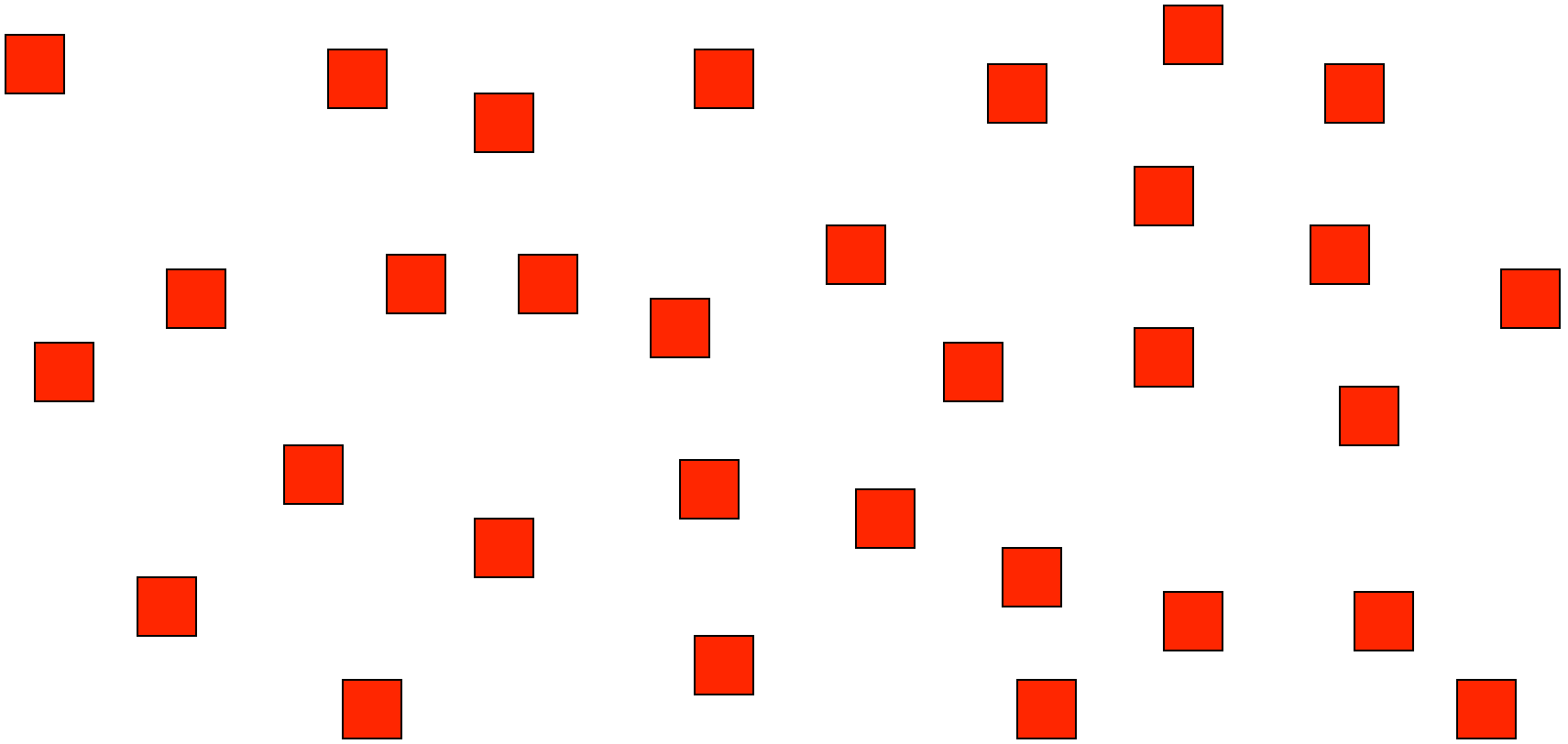
Is there a green square?



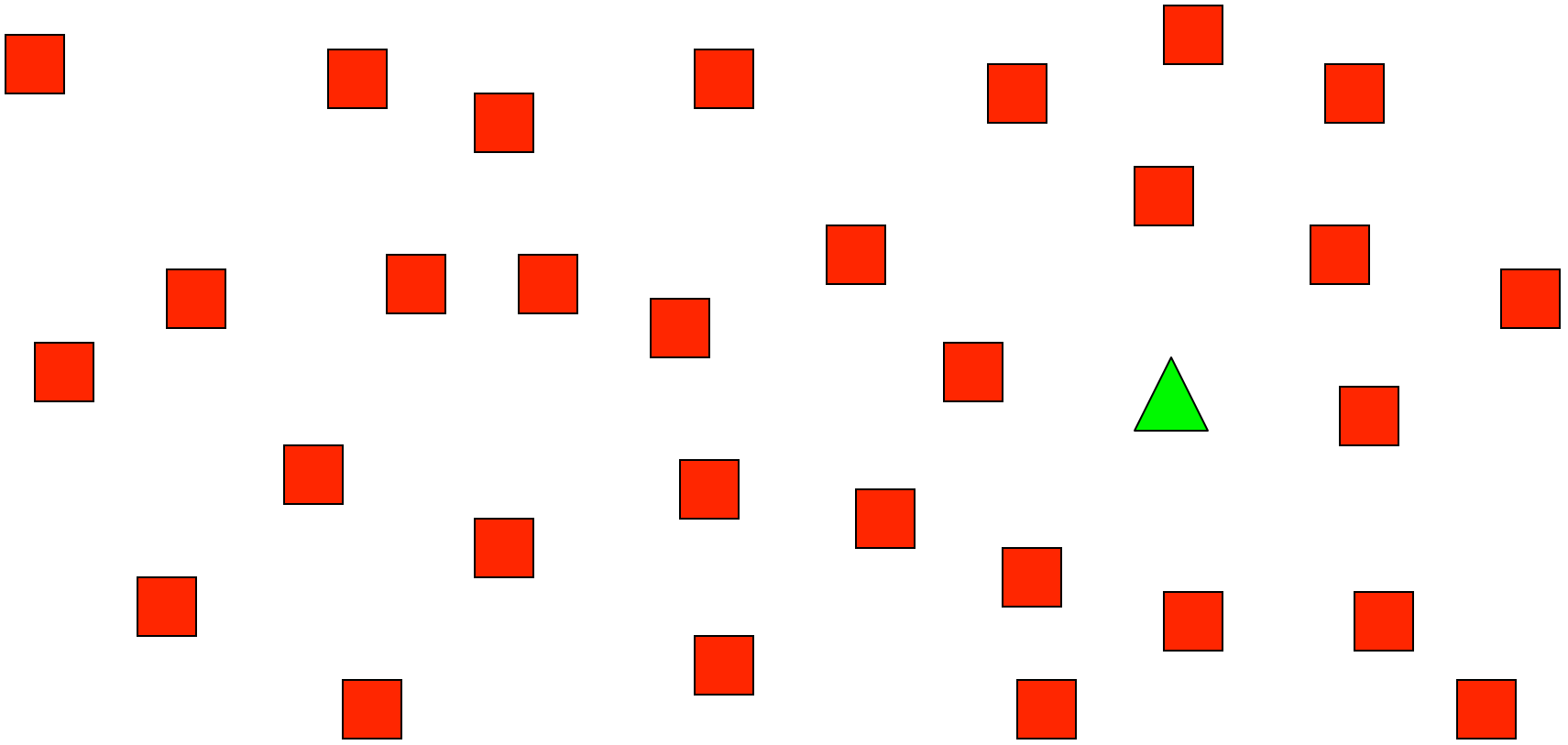
Is there a green square?



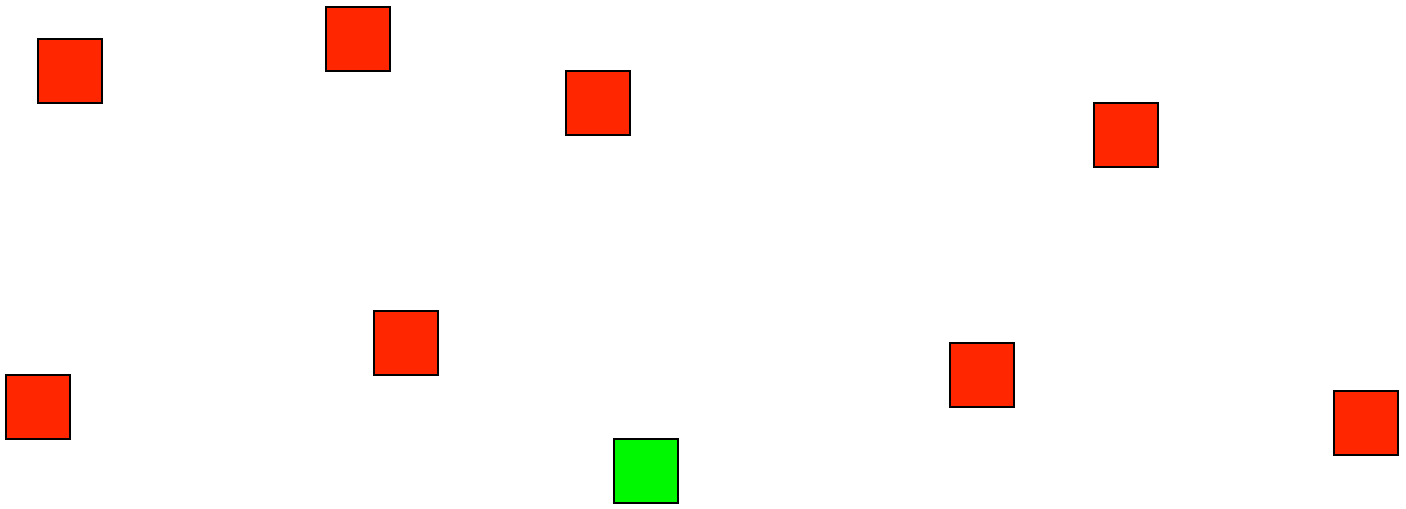
Is there a green square?



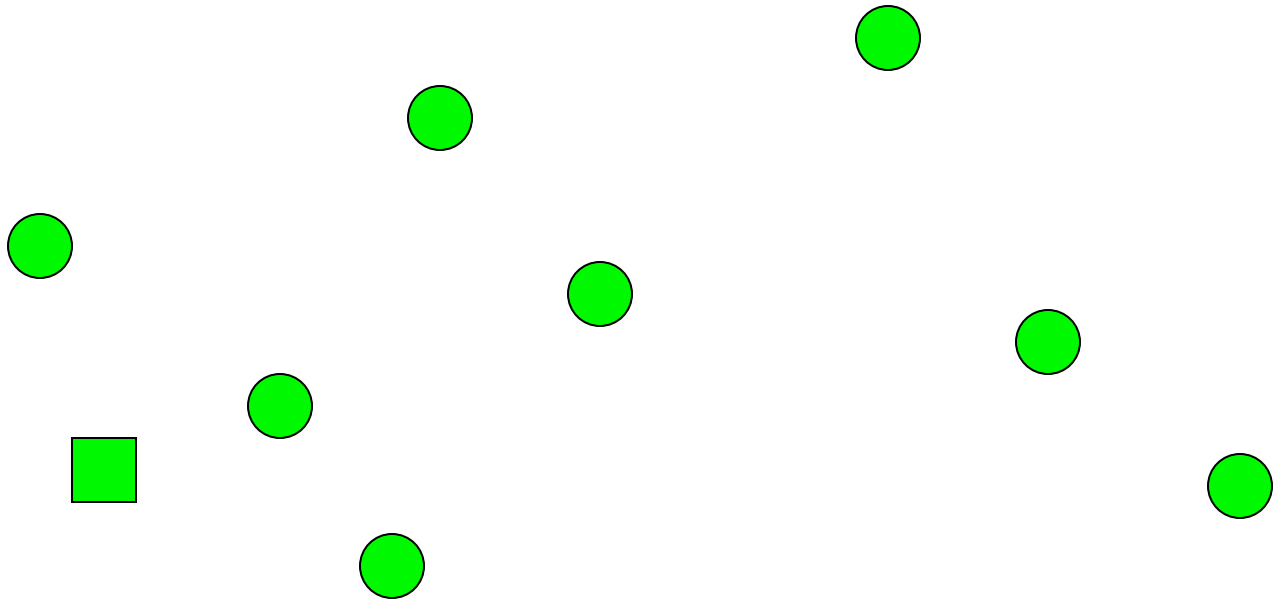
Is there a green square?



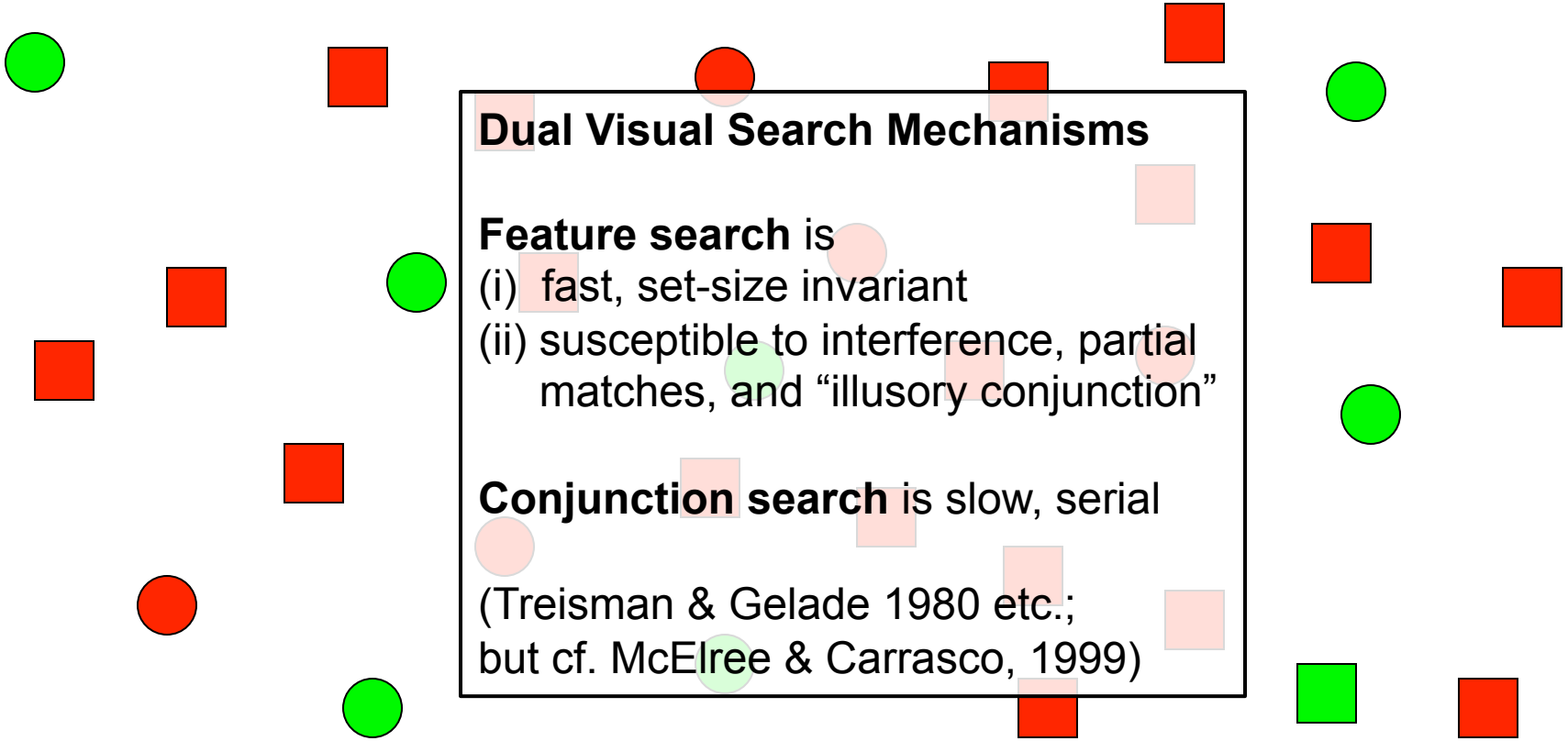
Is there a green square?



Is there a green square?



Is there a green square?



Dual Visual Search Mechanisms

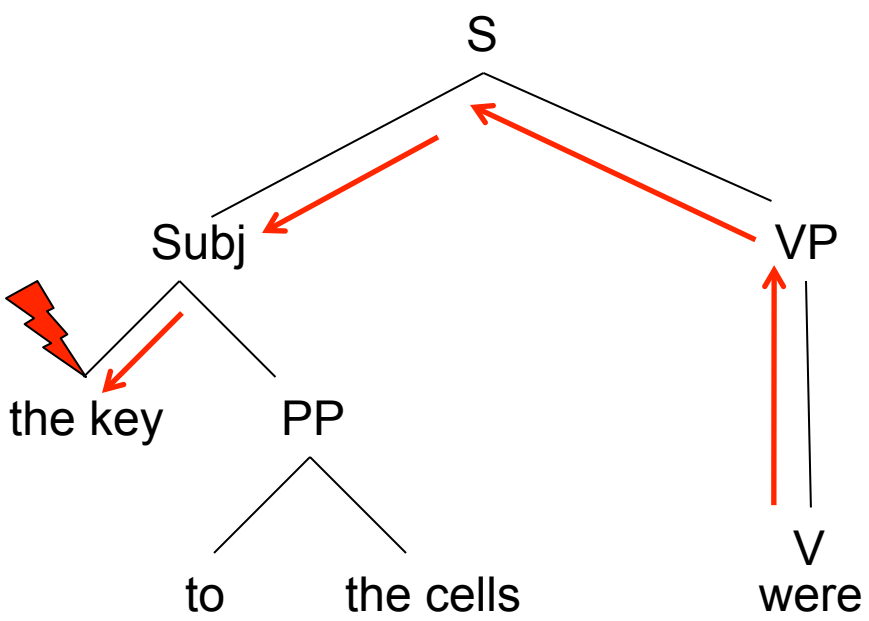
Feature search is
(i) fast, set-size invariant
(ii) susceptible to interference, partial matches, and "illusory conjunction"

Conjunction search is slow, serial

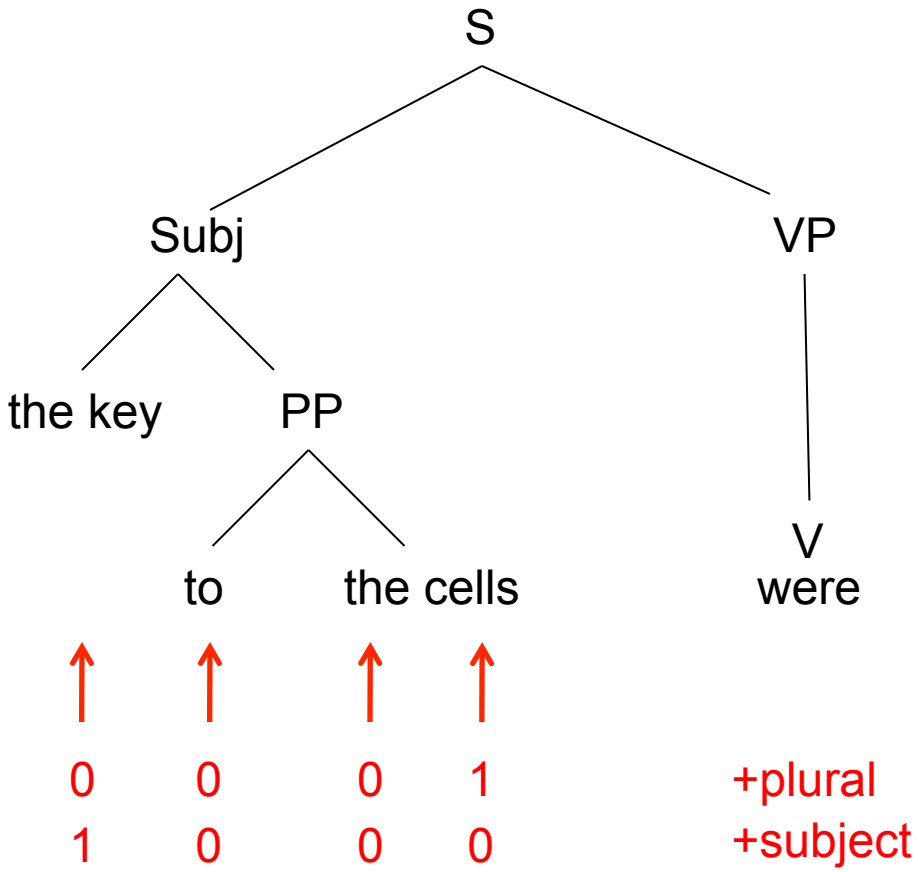
(Treisman & Gelade 1980 etc.;
but cf. McElree & Carrasco, 1999)

Two ways to search structures in memory

serial, structure-guided search



parallel, cue-guided (direct) access in *content-addressable memory*



structure-sensitive, avoids interference

slow, esp. for longer relations

susceptible to interference

fast, even for longer relations

McElree et al. 2003; Lewis et al. 2006; Martin & McElree 2008, 2009; Wagers, Lau, & Phillips, 2009

Content-addressable memory in language processing

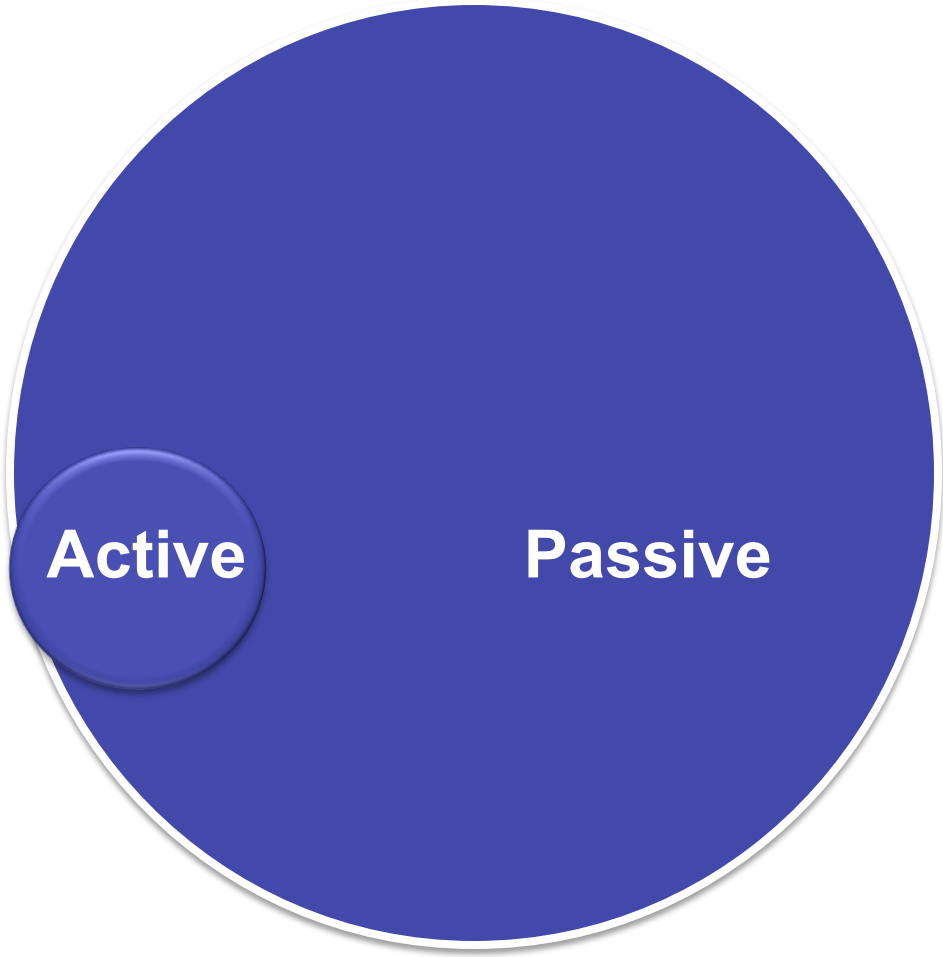
1. Timing

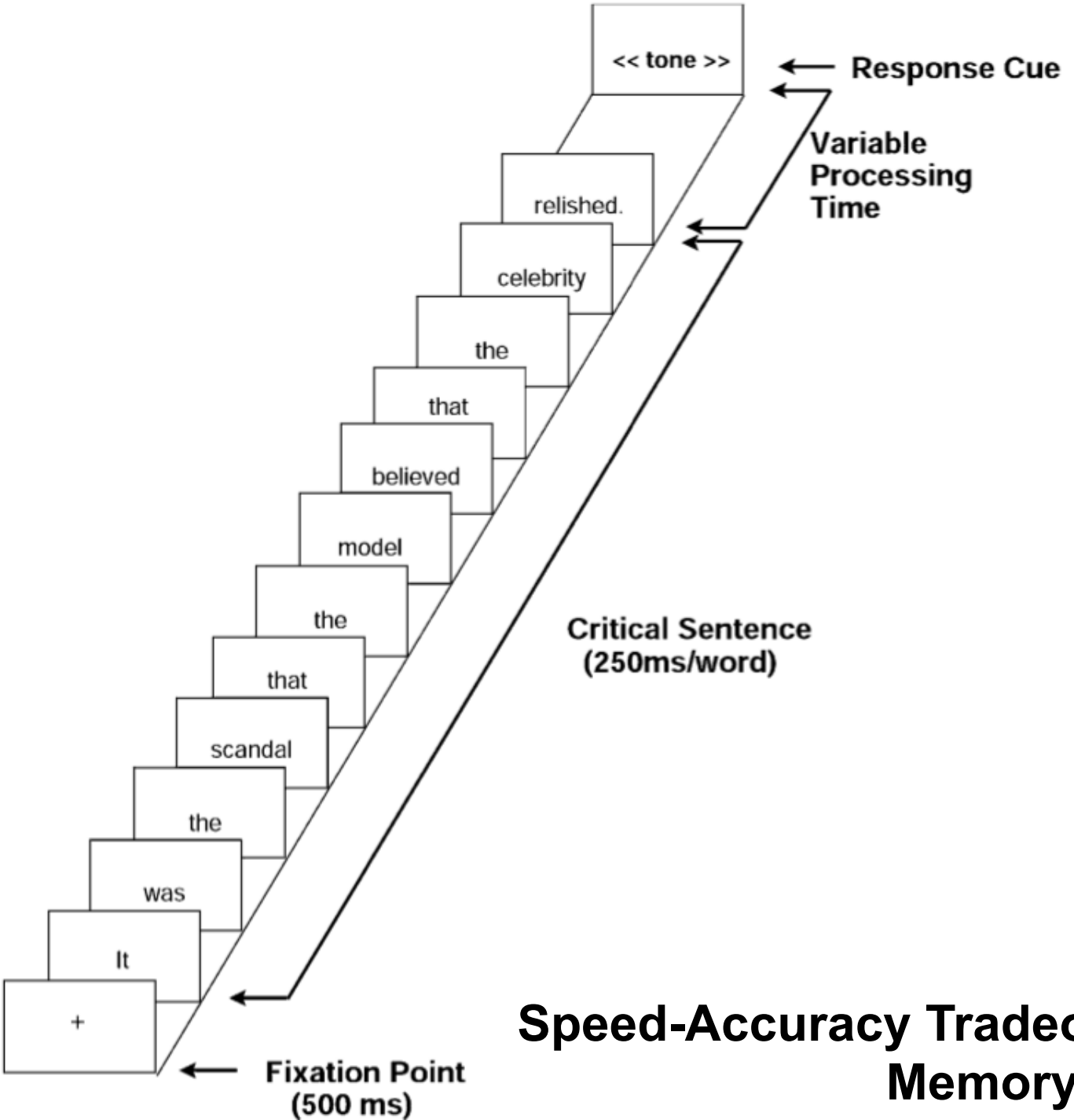
2. Facilitatory Interference

Bi-partite architecture of memory

Stringent limitations on the scope of information that can be concurrently processed

Broadbent 1958; Wickelgren et al., 1980; Garavan, 1998; Cowan, 2001; McElree, 2006; Verhaegen & Basak, 2007; Jonides et al., 2008

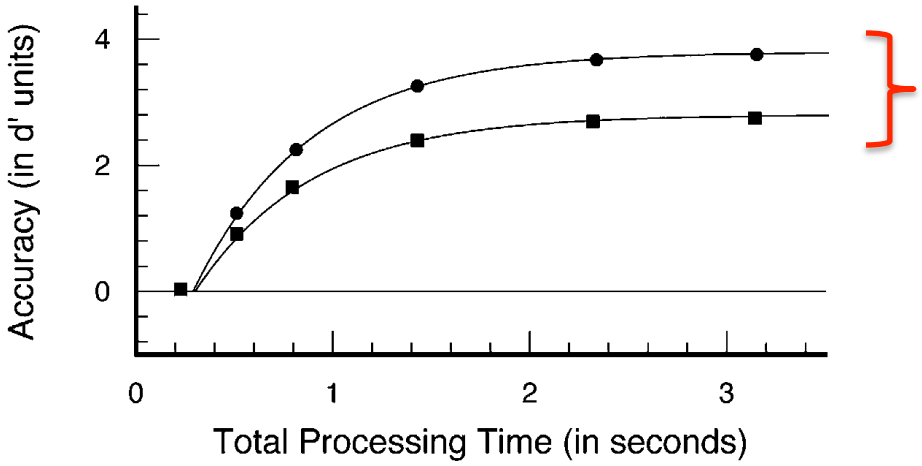




Speed-Accuracy Tradeoff (SAT) Memory Access

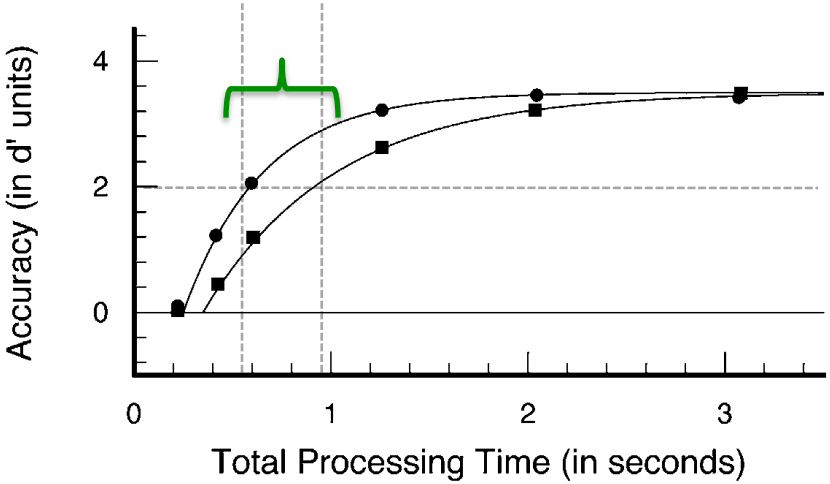
SAT: Possible Outcomes

Panel A



Asymptotic difference
Reflects the strength of the representation or the likelihood of completing a parse/process.

Panel B

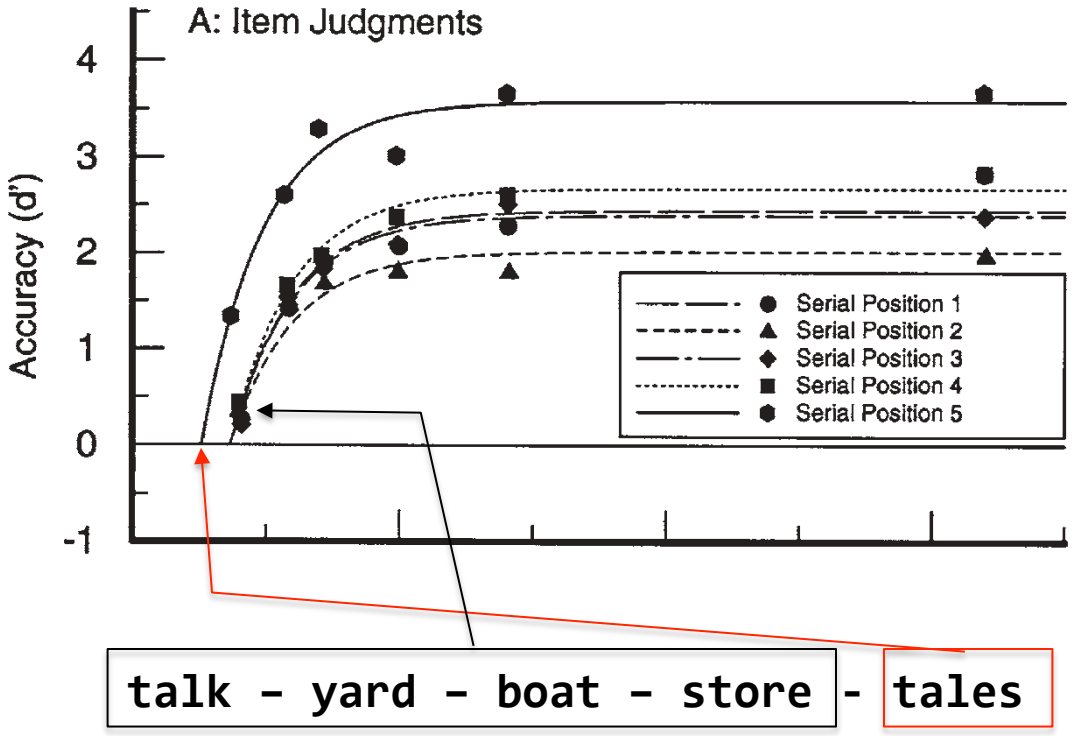


Rate/intercept difference
Reflects the speed of processing: how quickly information accumulates continuously, or the differences in an underlying discrete finishing time distribution.

Evidence from memory dynamics

Probe recognition – SAT response-signal task

Wickelgren, et al., 1980, McElree & Dosher, 1989



SLOOOOOWWWW ***FAST***

ACADEMIC PRESS

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Journal of Memory and Language 48 (2003) 67–91

Journal of Memory and Language

www.elsevier.com/locate/jml

Memory structures that subserve sentence comprehension

Brian McElree,* Stephani Foraker, and Lisbeth Dyer

Department of Psychology, New York University, 6 Washington Place, Room 859, New York, NY 10003, USA

Received 2 August 2001; revision received 24 May 2002

Abstract

Measures of the speed and accuracy of processing sentences with nonadjacent dependencies derived from the response-signal speed-accuracy tradeoff procedure were used to examine the nature of the memory system that underlies sentence comprehension. Three experiments with different sentence structures demonstrated that the accuracy of processing a dependency decreased as more material was interpolated between nonadjacent constituents. However, processing speed was unaffected by the amount of interpolated material, indicating that memory representations for previously processed constituents can be accessed directly. These results suggest that a content-addressable memory system mediates sentence comprehension, in which syntactic and semantic information provide direct access to memory representations without the need to search through extraneous representations. Notably, content-addressability appears to underlie the interpretation of sentence structures that also require the recovery of order information, a type of operation that has been shown to necessitate a slow search process in list-learning experiments (McElree, 2001; McElree & Dosher, 1993).

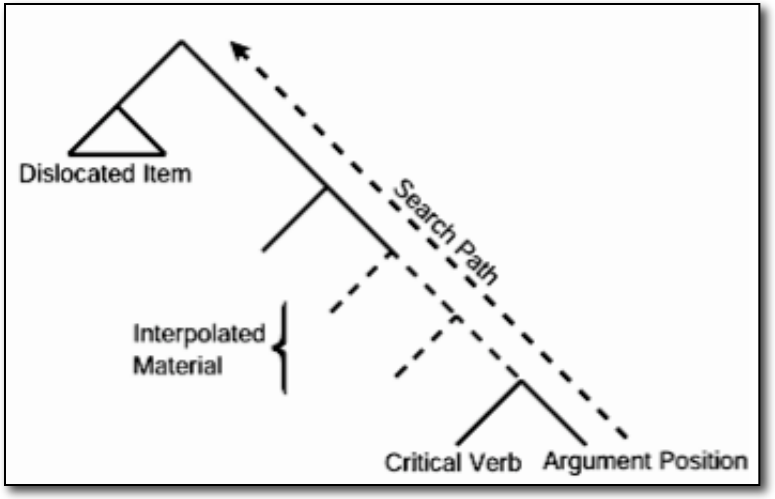
Keywords: Sentence comprehension; Working memory; Unbounded dependencies; Memory retrieval; Speed-accuracy tradeoff

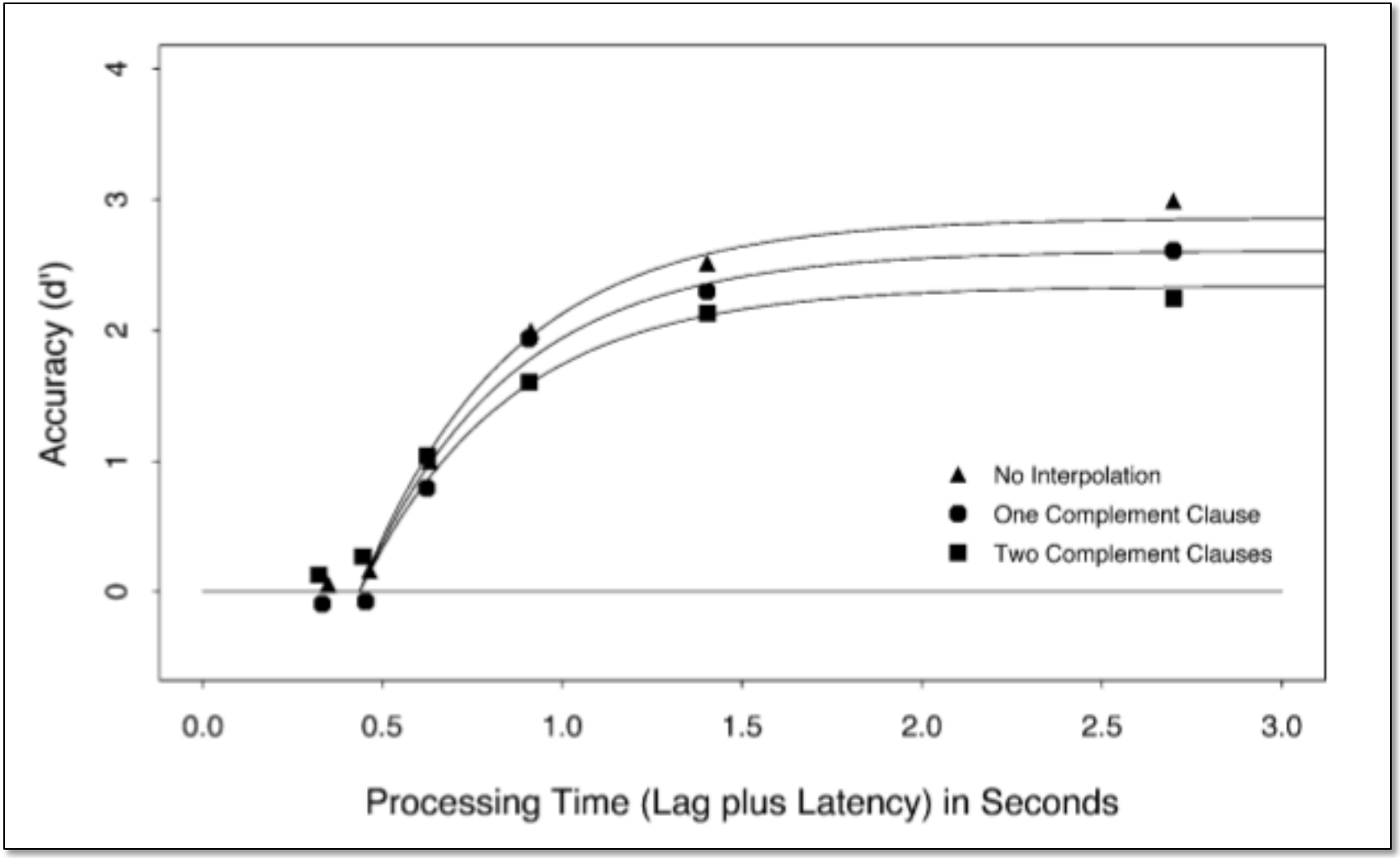
McElree et al. 2003

Constructions used in Experiment 1

Construction type	Acceptability	Example
No interpolation	Acceptable	T1. It was the scandal that the celebrity relished
No interpolation	Unacceptable	T2. It was the scandal that the celebrity panicked
One interpolated clause	Acceptable	T3. It was the scandal that the model believed that the celebrity relished
One interpolated clause	Unacceptable	T4. It was the scandal that the model believed that the celebrity panicked
Two interpolated clauses	Acceptable	T5. It was the scandal that the model believed that the journalist reported that the celebrity relished
Two interpolated clauses	Unacceptable	T6. It was the scandal that the model believed that the journalist reported that the celebrity panicked
Controls for relative clause processing		
One interpolated clause	Unacceptable	T7. It was the scandal that the model amused that the celebrity relished
One interpolated clause	Unacceptable	T8. It was the scandal that the model believed that the journalist amused that the celebrity relished
Additional controls		
No interpolation	Acceptable	T9. The scandal panicked the celebrity
No interpolation	Unacceptable	T10. The scandal relished the celebrity
One interpolated clause	Acceptable	T11. The model believed that the scandal panicked the celebrity
One interpolated clause	Unacceptable	T12. The model believed that the scandal relished the celebrity
One interpolated clause	Unacceptable	T13. The model ambled that the scandal panicked the celebrity
Two interpolated clauses	Acceptable	T14. The model believed that the journalist reported that the scandal panicked the celebrity
Two interpolated clauses	Unacceptable	T15. The model believed that the journalist ambled that the scandal panicked the celebrity
Two interpolated clauses	Unacceptable	T16. The model believed that the journalist reported that the scandal relished the celebrity

Acceptable	T1. It was the scandal that the celebrity relished
Unacceptable	T2. It was the scandal that the celebrity panicked
Acceptable	T3. It was the scandal that the model believed that the celebrity relished
Unacceptable	T4. It was the scandal that the model believed that the celebrity panicked
Acceptable	T5. It was the scandal that the model believed that the journalist reported that the celebrity relished
Unacceptable	T6. It was the scandal that the model believed that the journalist reported that the celebrity panicked





Distance affects asymptote, but not temporal dynamics

Interpretation: whole sentence accessed in parallel – no serial search

Concern: wh-constructions are not a good test of distance effects

Agreement

The New York Times

"And in the absence of large-scale policy differences between the two candidates, the personal nature of their exchanges are more likely to result in lasting damage." [4/9/08]

The Washington Post

"Republicans privately acknowledge this, arguing that in the hands of a more popular politician, the ideas that Cheney are putting forward could find fertile ground with the American people." [5/21/09]

Agreement Illusions

Not only do we produce agreement errors – we generally fail to notice them

"**The key** to the cells unsurprisingly **were** rusty …"

"**The key** to the cell unsurprisingly **were** rusty …"

It's not simply 'proximity concord':

"The musicians who **the reviewer praise** so highly …"

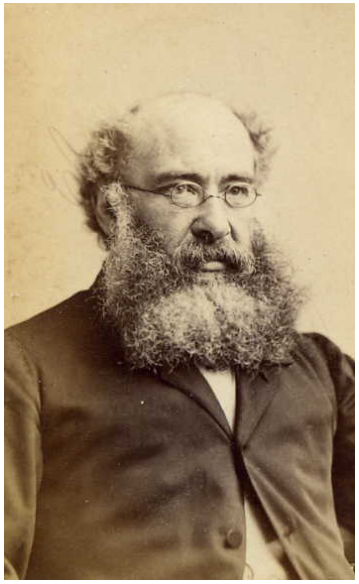
"The musician who **the reviewer praise** so highly …"

And it is selective – plurals create illusions, singulars don't

"**The keys** to the cell unsurprisingly **was** rusty …"

And it happens all the time …

(Bock & Miller 1991; Pearlmutter et al. 1999; Deevy et al. 1998; Staub 2009; Wagers, Lau, & Phillips 2009; Eberhard et al. 2005)

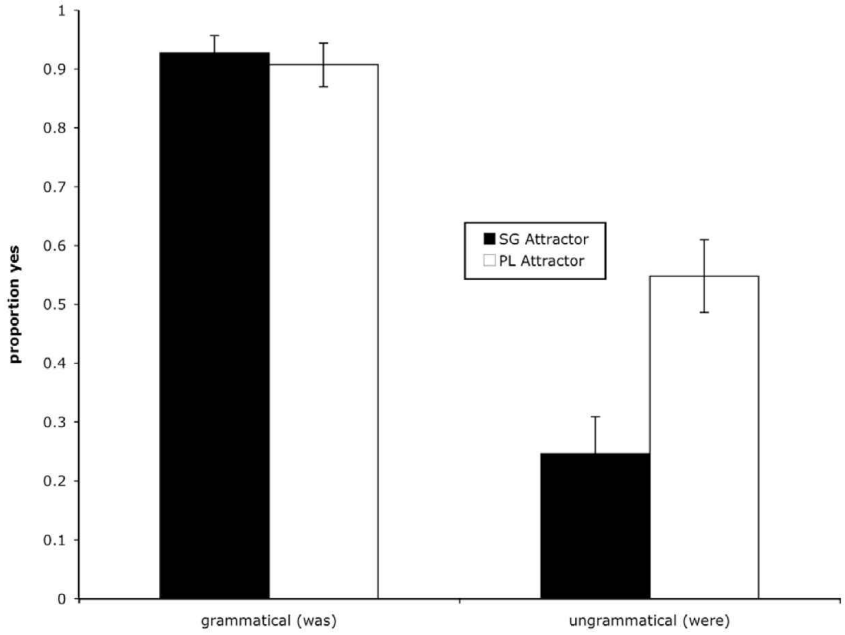


“Rapid writing will no doubt give rise to inaccuracy. … A singular nominative will be disgraced by a plural verb, because other pluralities have intervened and have tempted the ear into plural tendencies. I am ready to declare that, with much training, I have been unable to avoid them.” (Anthony Trollope, 1883)

What causes agreement attraction?

1. Encoding: hallucinating plural subject

2. Access: misretrieving irrelevant noun



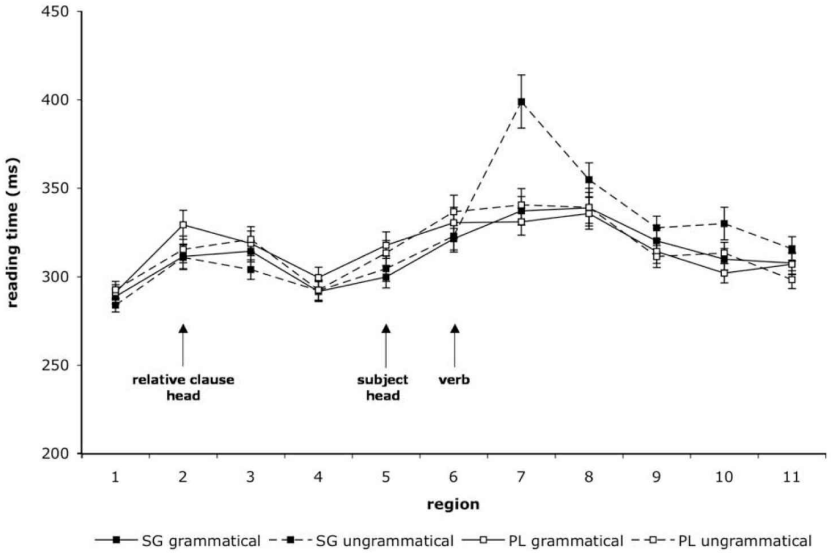
Evidence: Grammatical Asymmetry

Illusions of acceptability

The key to the cabinet were
The key to the cabinets were

No illusions of unacceptability

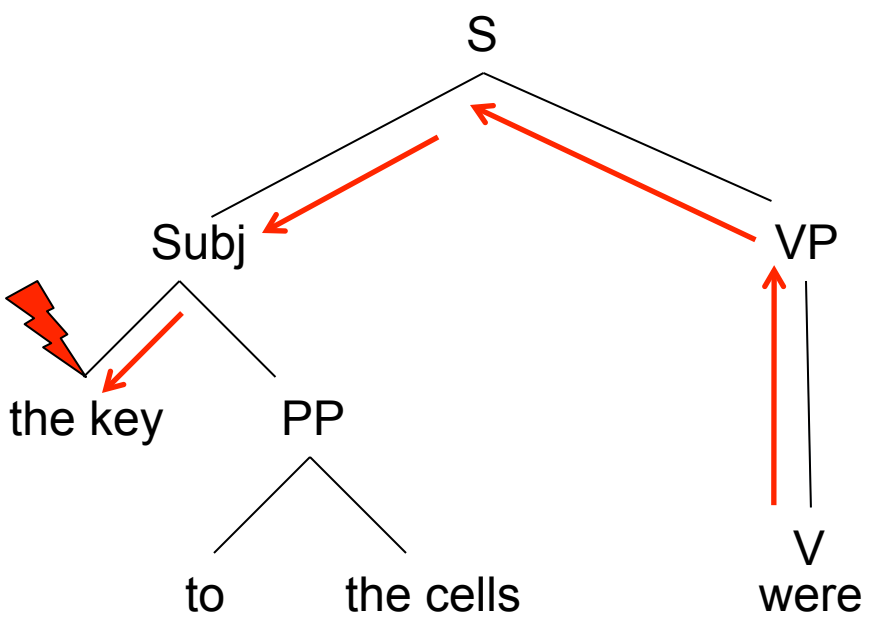
The key to the cabinet was
The key to the cabinets was



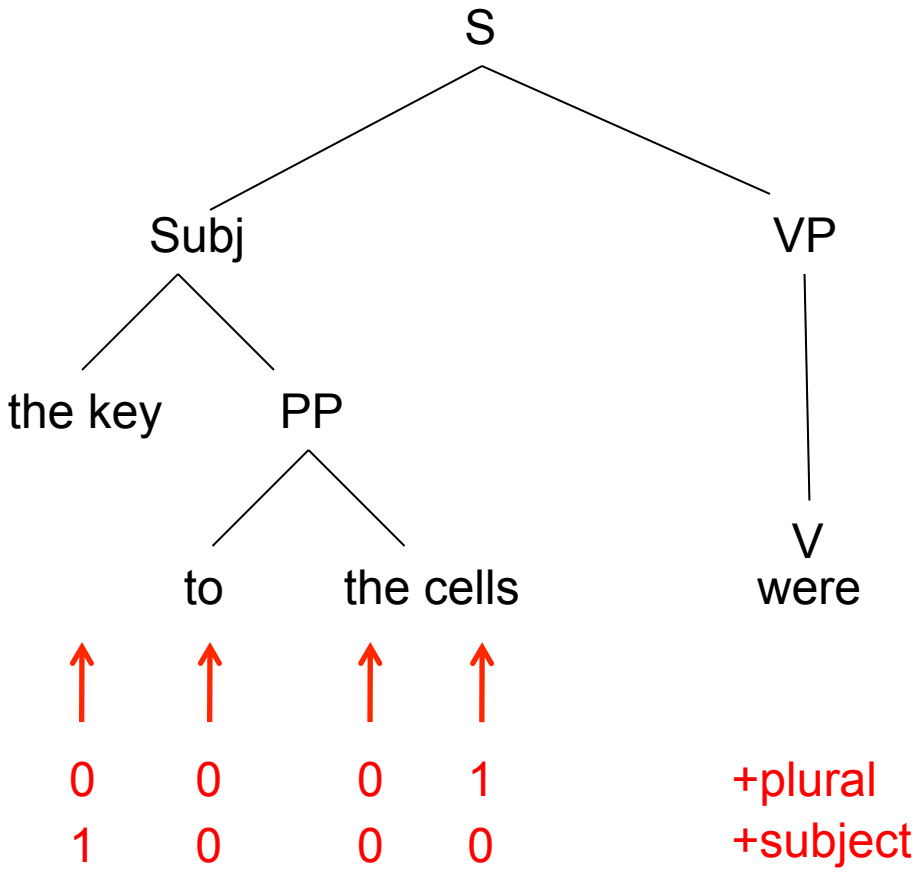
Wagers, Lau, & Phillips 2009, *J Mem Lang*

Two ways to search structures in memory

serial, structure-guided search



parallel, cue-guided (direct) access in *content-addressable memory*



structure-sensitive, avoids interference

slow, esp. for longer relations

susceptible to interference

fast, even for longer relations

McElree et al. 2003; Lewis et al. 2006; Martin & McElree 2008, 2009; Wagers, Lau, & Phillips, 2009

Opinion TRENDS in Cognitive Sciences Vol.10 No.10

Computational principles of working memory in sentence comprehension

Richard L. Lewis[1], Shravan Vasishth[2] and Julie A. Van Dyke[3]

[1] Department of Psychology, University of Michigan, 530 Church St, Ann Arbor, MI 48109, USA
[2] Institute for Linguistics, University of Potsdam, Postfach 601553, D-14415 Potsdam, Germany
[3] Haskins Laboratories, 300 George Street, New Haven, CT 06511, USA

Understanding a sentence requires a working memory of the partial products of comprehension, so that linguistic relations between temporally distal parts of the sentence can be rapidly computed. We describe an emerging theoretical framework for this working memory system that incorporates several independently motivated principles of memory: a sharply limited attentional focus, rapid retrieval of item (but not order) information subject to interference from similar items, and activation decay (forgetting over time). A computational model embodying these principles provides an explanation of the functional capacities and severe limitations of human processing, as well as accounts of reading times. The broad implication is that the detailed nature of crosslinguistic sentence processing emerges from the interaction of general principles of human memory with the specialized task of language comprehension.

metrics to characterize the memory load associated with linguistic structures [2,3] but this work has not made significant contact with theories of memory in cognitive psychology [4], and has not yielded models of the underlying working memory architecture.

Here, we present an emerging theoretical framework that seeks to bridge this gap and directly answer the question above. The basic framework claims that language processing, although it might operate on specialized representations, is nevertheless subject to general processing principles and constraints that govern other domains of memory [5–7]. The application of these principles to sentence processing results in a surprising breadth of explanation and a range of detailed predictions, which we review here.

This approach builds on several important recent trends:

- Increased empirical atten… phenomena that reveal constr… independently of ambiguity resolution [8–14]

Lewis et al. 2006

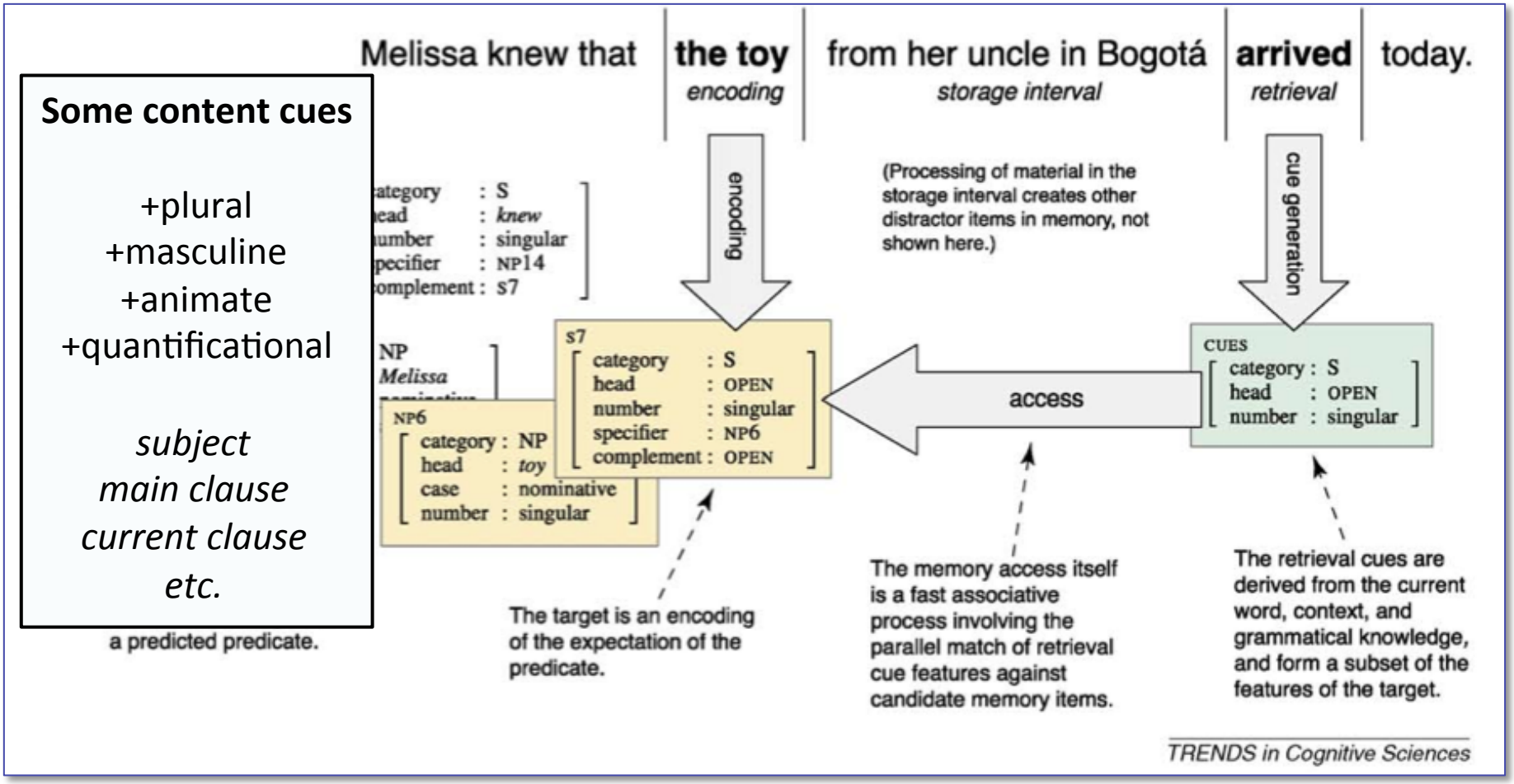
Memory-access Model: ACT-R

Key features

(i) Parallel-access → time-constant retrieval

(ii) Content-addressable → susceptible to partial-match interference

(iii) Limited-size buffers → ~ restricted focus of attention; const. shunting



(Lewis, Vasishth, & van Dyke 2006, *TICS*; cf. McElree 2006)

Same Memory – Different Access

Subject-Verb Agreement

The diva [that accompanied the **harpist** on stage] clearly **was** flawless …
The diva [that accompanied the **harpists** on stage] clearly **was** flawless …

The diva [that accompanied the **harpist** on stage] clearly **were** flawless …
The diva [that accompanied the **harpists** on stage] clearly **were** flawless …

illusion

Subject-Reflexive Agreement

The diva [that accompanied the **harpist** on stage] clearly presented **herself** …
The diva [that accompanied the **harpists** on stage] clearly presented **herself** …

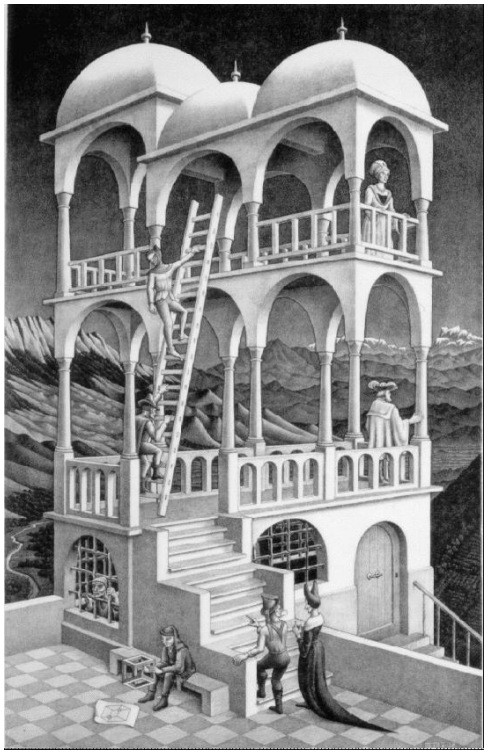
The diva [that accompanied the **harpist** on stage] clearly presented **themselves** …
The diva [that accompanied the **harpists** on stage] clearly presented **themselves** …

no illusion

Both processes require access to same element
-- the subject of the same clause.

Dillon, Mishler, Sloggett, & Phillips, 2013

Grammatical Illusions



- **Agreement**
  Clifton, Frazier & Deevy, 1999; Wagers, Lau, & Phillips 2009
- **Negative Polarity**
  Vasishth et al. 2008; Xiang et al. 2009; Parker & Phillips 2013
- **Comparatives**
  Wellwood et al. in prep
- **Case**
  Bader & Bayer 2006; Sloggett 2013
- **Thematic assignment**
  Ferreira et al. 2003; Kim & Osterhout 2005; Kuperberg 2007
  BUT Chow & Phillips 2013
- **Anaphora: Principle B** (disputed)
  Badecker & Straub 2002; Lee & Williams 2006
  BUT Nicol & Swinney 1989; Clifton et al. 1997; Runner et al. 2006; Lewis et al. 2012

On-line Grammatical Sensitivity

- **Anaphora: Principle C**
  English: Kazanina et al. 2007; Japanese: Aoshima et al. 2009;
  Russian: Kazanina & Phillips 2010
  English-crossover: Kush, Lidz, & Phillips 2013
- **Anaphora: Principle A**
  Nicol & Swinney 1989; Clifton et al., 1998; Badecker & Straub 2002; Sturt 2003; Xiang et al. 2009; Dillon et al. 2013
- **Islands**
  Subjects: Stowe 1986; Bourdages 1992; Traxler & Pickering 1996; Omaki & Schulz 2010; Rel. Cl.: McElree & Griffith 1998; CSC: Wagers & Phillips 2009; wh-islands: Neville et al. 1991; Japanese: Yoshida et al. 2004; Parasitic gaps: Phillips 2006
- **Sluicing**
  Yoshida et al. 2011
- **Agreement** (selective)
  Wagers, Lau, & Phillips 2009
- **Case** (selective)
  Bader & Bayer 2006
- **Anaphora: Principle B** (disputed)
  Nicol & Swinney 1989; Clifton et al. 1997; Runner et al. 2006; Lewis et al. 2012
  BUT: Badecker & Straub 2002; Lee & Williams 2006

Generalizations

- Successes imply richly structured mental encodings

  (yes, we already knew they exist, now we know they're created quickly)

- If the representations are intact, then the failures must reflect

  – Intact grammar + noisy access mechanisms (cue-based retrieval)
  – … and some other sources that don't concern us here (cf. S. Lewis & Phillips 2015)
  – We do not want to simply allow different phenomena to work differently – learning!

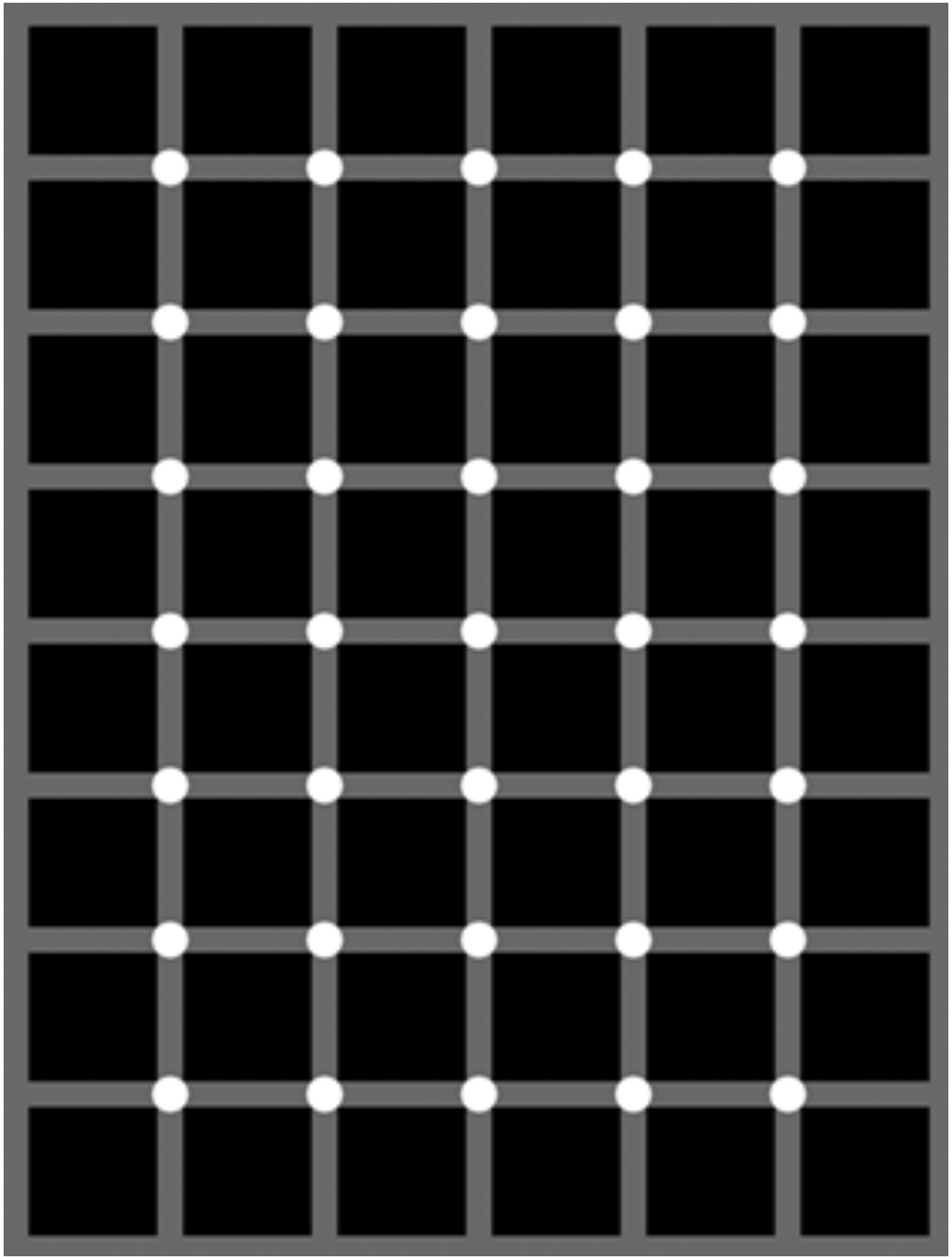
- So why are the failures selective? (cf. Phillips, Wagers, & Lau 2011)

  – Forwards vs. backwards looking processes ('forewarned is forearmed')
  – Dependency-wise differences in access mechanisms or use of features (e.g., agreement vs. anaphora)
  – Availability of relational structural notions in access mechanisms
    → example of *c-command* (we'll get back to that in a moment)

Dan Parker

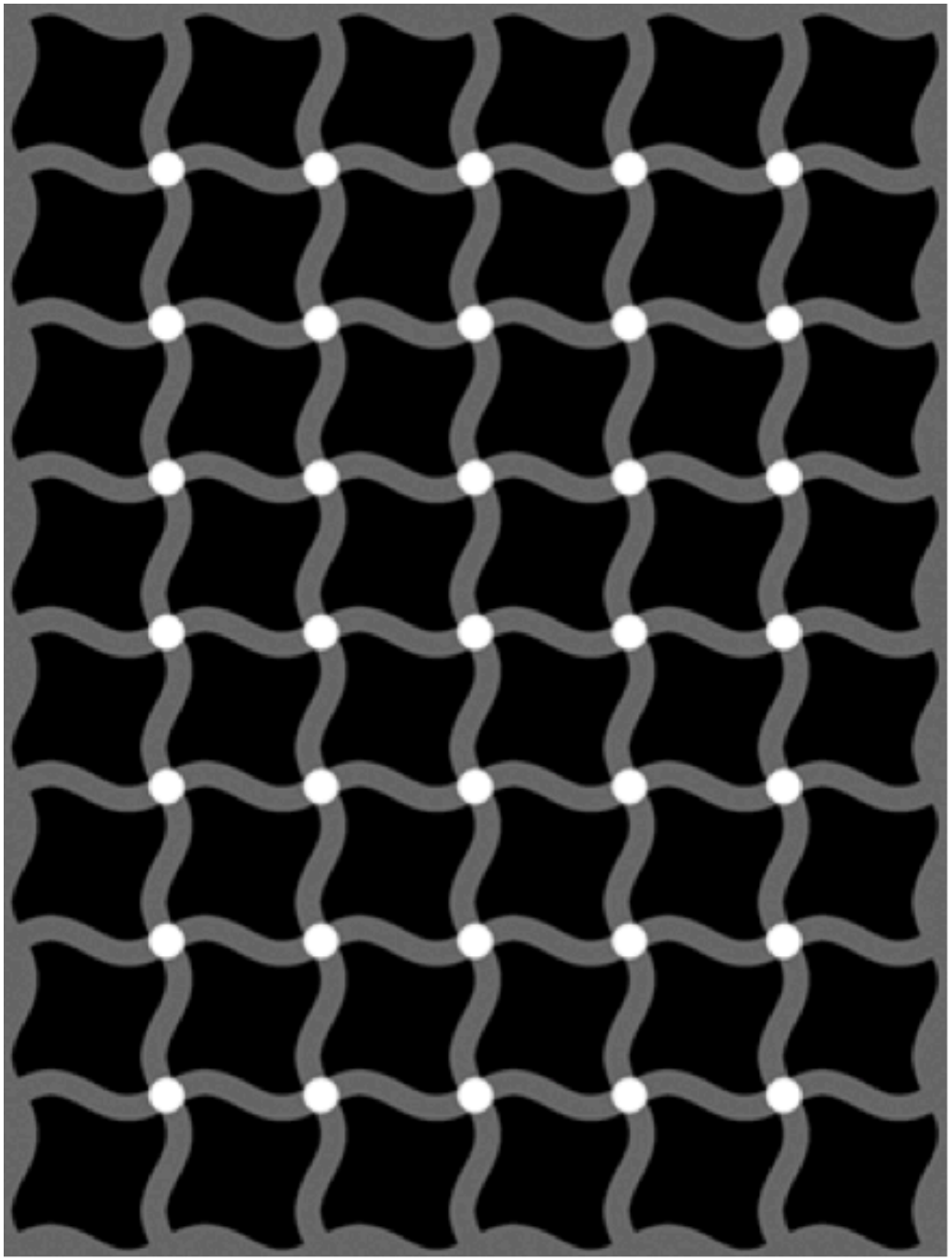
Surprise #1

GRAIN SIZE OF SELECTIVE FALLIBILITY



scintillating grid illusion

Lingelbach 1994



scintillating grid illusion

… disappears

Geier et al 2008

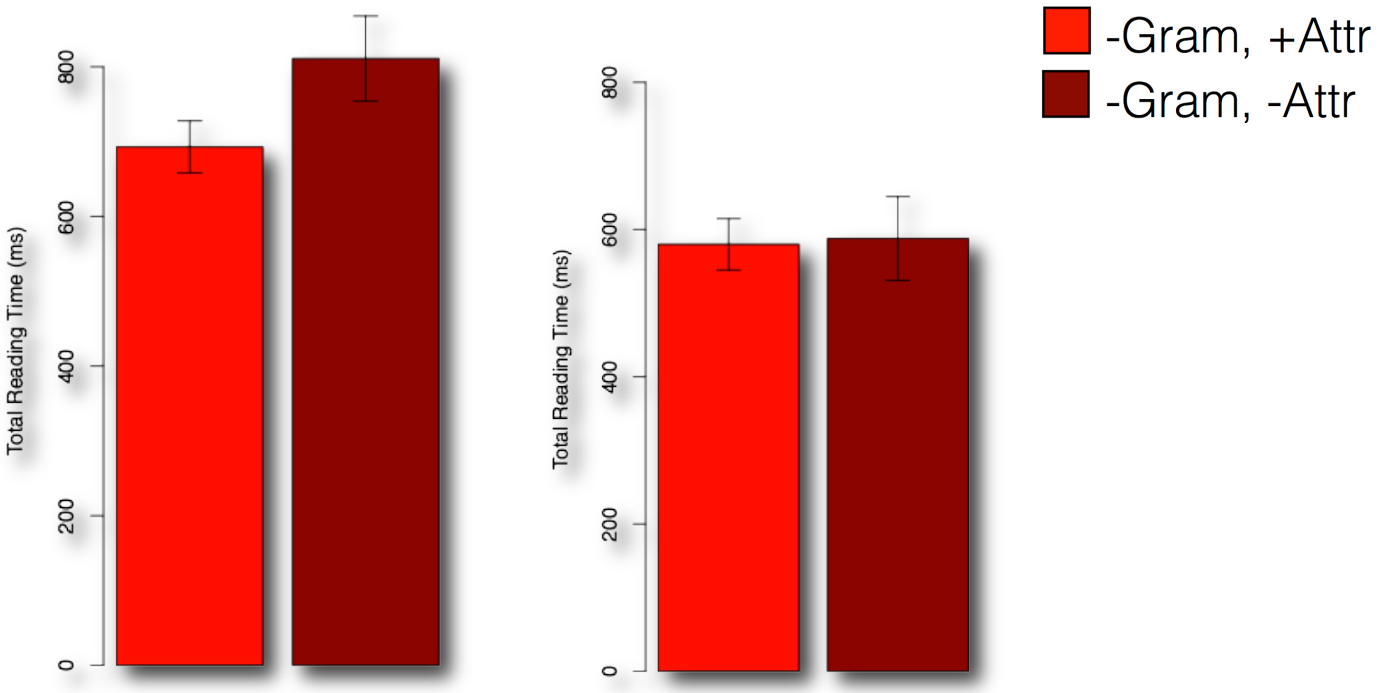
Agreement attraction

*<u>The manager</u> [who oversaw <u>the employee(s)</u>] **were** dishonest

(eye-tracking data from Dillon, Mishler, Sloggett & Phillips 2013)

Reflexive non-attraction

*<u>The manager</u> [who oversaw <u>the employee(s)</u>] doubted **themselves**



(eye-tracking data from Dillon, Mishler, Sloggett & Phillips 2013)

Gender non-attraction

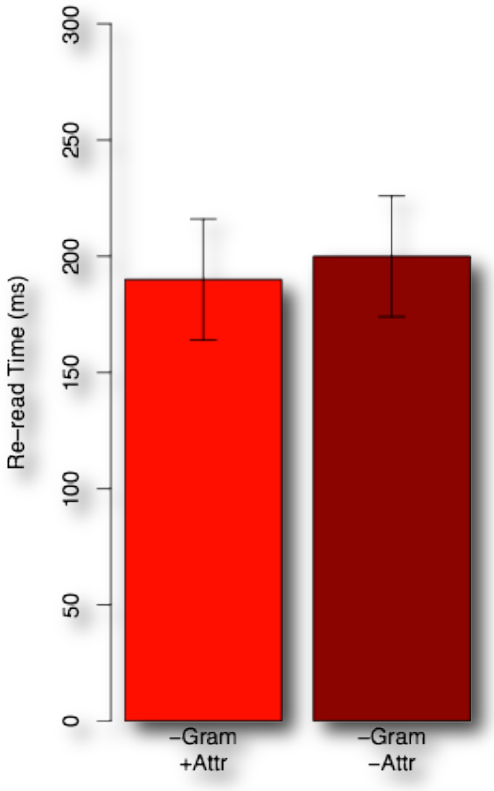
- Reflexive and antecedent must also agree in gender

*He/She said that [the nurse pricked **himself**]

Consistent finding:
no illusions-of-acceptability w/ reflexives
Sturt 2003; Xiang et al. 2009; Dillon et al. 2013; Kush et al. 2012; Nicol & Swinney, 1989; Clifton et al. 1999; Kennison & Trofe, 2003; Patil et al. 2011, etc.

Possibility #1: serial search (Dillon in press)
Possibility #2: only syntactic cues (Dillon et al. 2013)

(Sturt 2003, expt. 1)

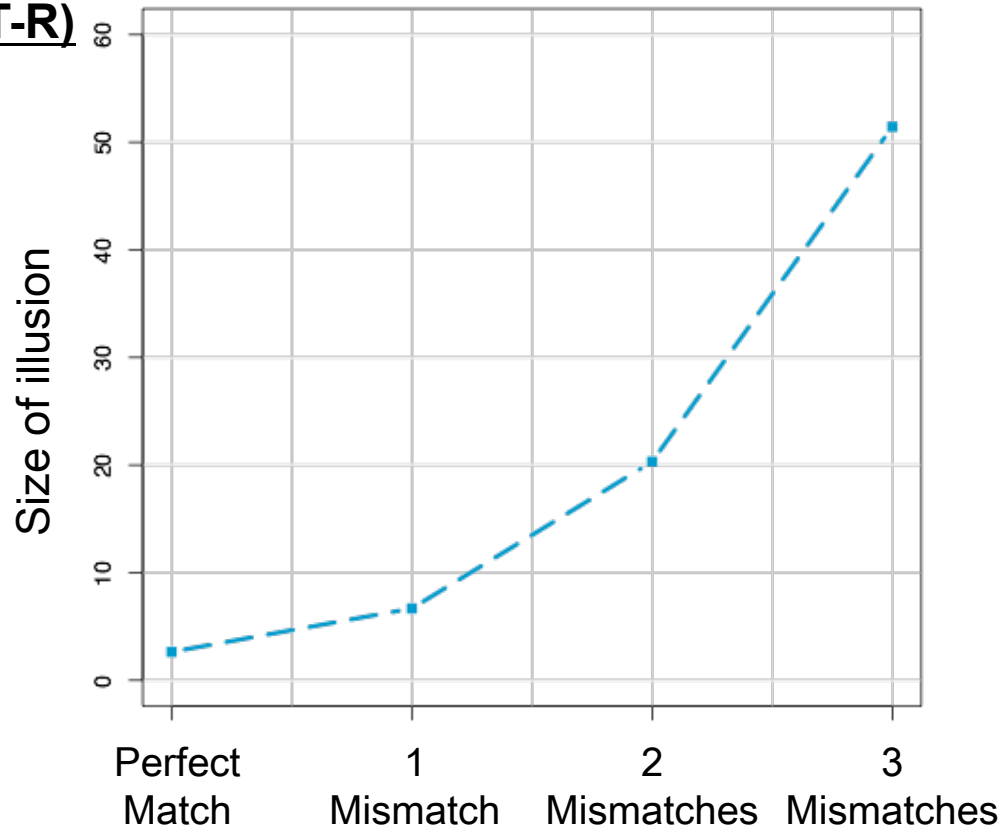


Stronger mismatches

Previous studies of reflexives limited to contexts involving a single feature mismatch, e.g., number or gender

<u>Surprising modeling prediction (ACT-R)</u>

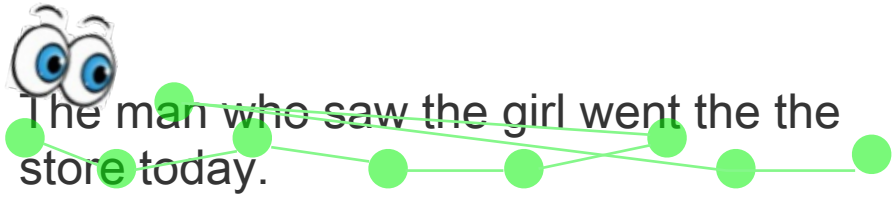
Multiple feature mismatches
➜exponential increase in presence of illusion



Experiments 1a-c

Comparison: contexts involving 1- vs. 2-feature mismatches

Method: eye-tracking while reading



Experiment 1a: gender + number

Experiment 1b: gender + animacy

Experiment 1c: number + animacy

1-feature mismatch: gender mismatch (items adapted from Sturt 2003)

Grammatical with lure

The **librarian** said [that the **schoolgirl** reminded **herself** about the assignment].

Grammatical without lure

The **janitor** said [that the **schoolgirl** reminded **herself** about the assignment].

Ungrammatical with lure

*The **librarian** said [that the **schoolboy** reminded **herself** about the assignment].

Ungrammatical without lure

*The **janitor** said [that the **schoolboy** reminded **herself** about the assignment].

2-feature mismatches: gender + number mismatch

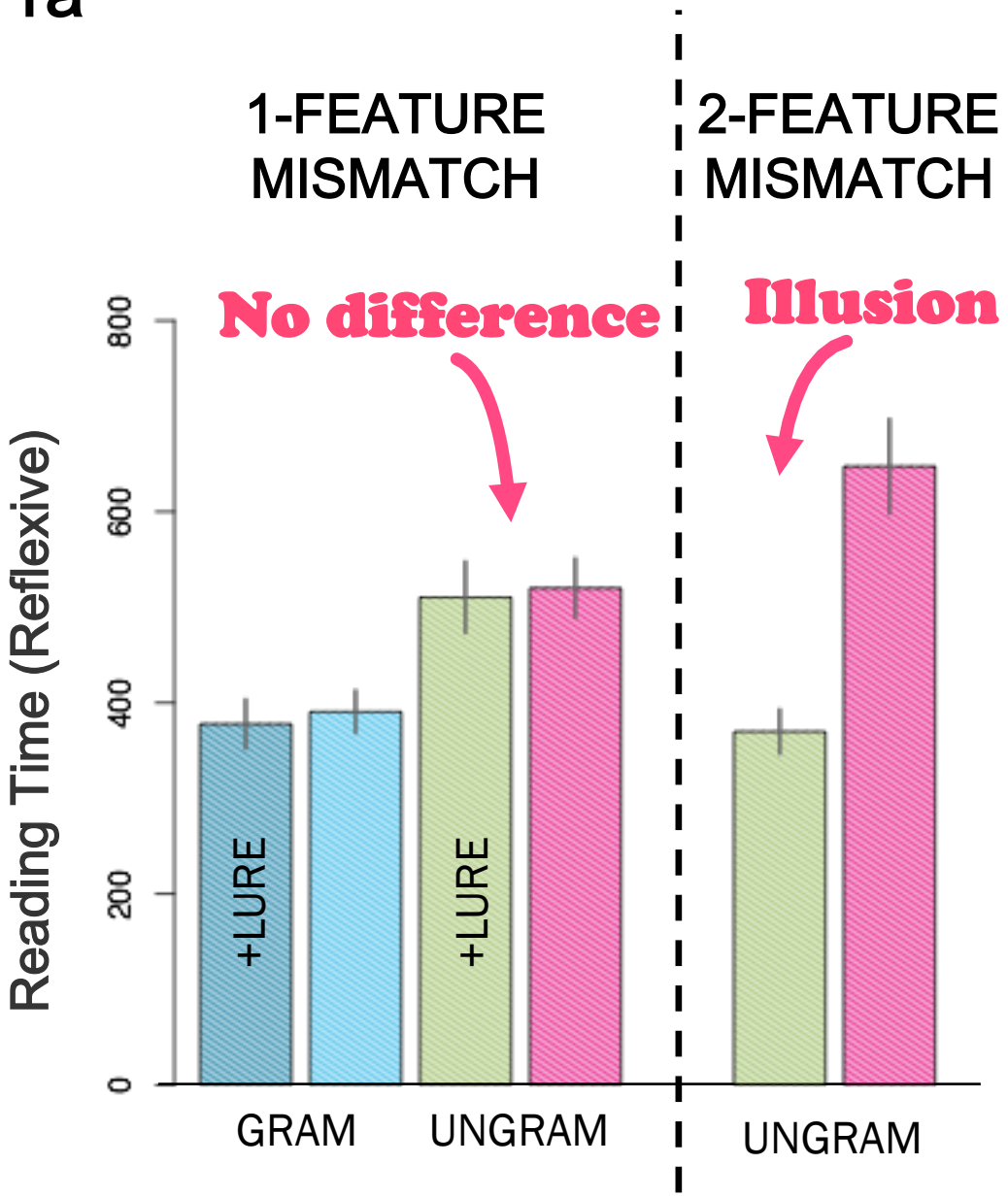
Ungrammatical with lure

*The **librarian** said [that the **schoolboys** reminded **herself** about the assignment].

Ungrammatical without lure

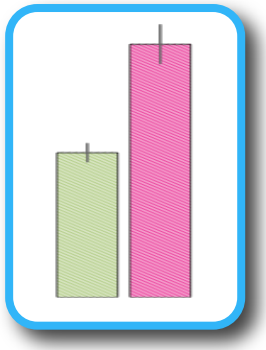
*The **janitor** said [that the **schoolboys** reminded **herself** about the assignment].

Experiment 1a

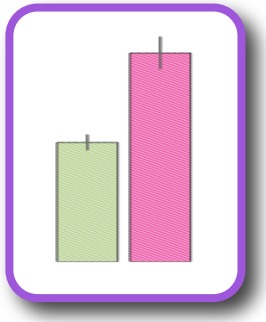


Experiments 1a-c

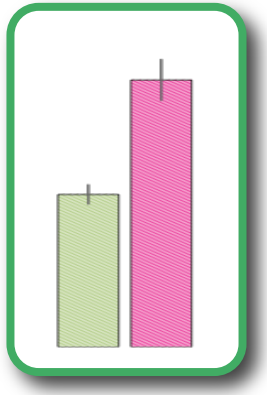
Experiment 1a: gender + number



Experiment 1b: gender + animacy



Experiment 1c: number + animacy



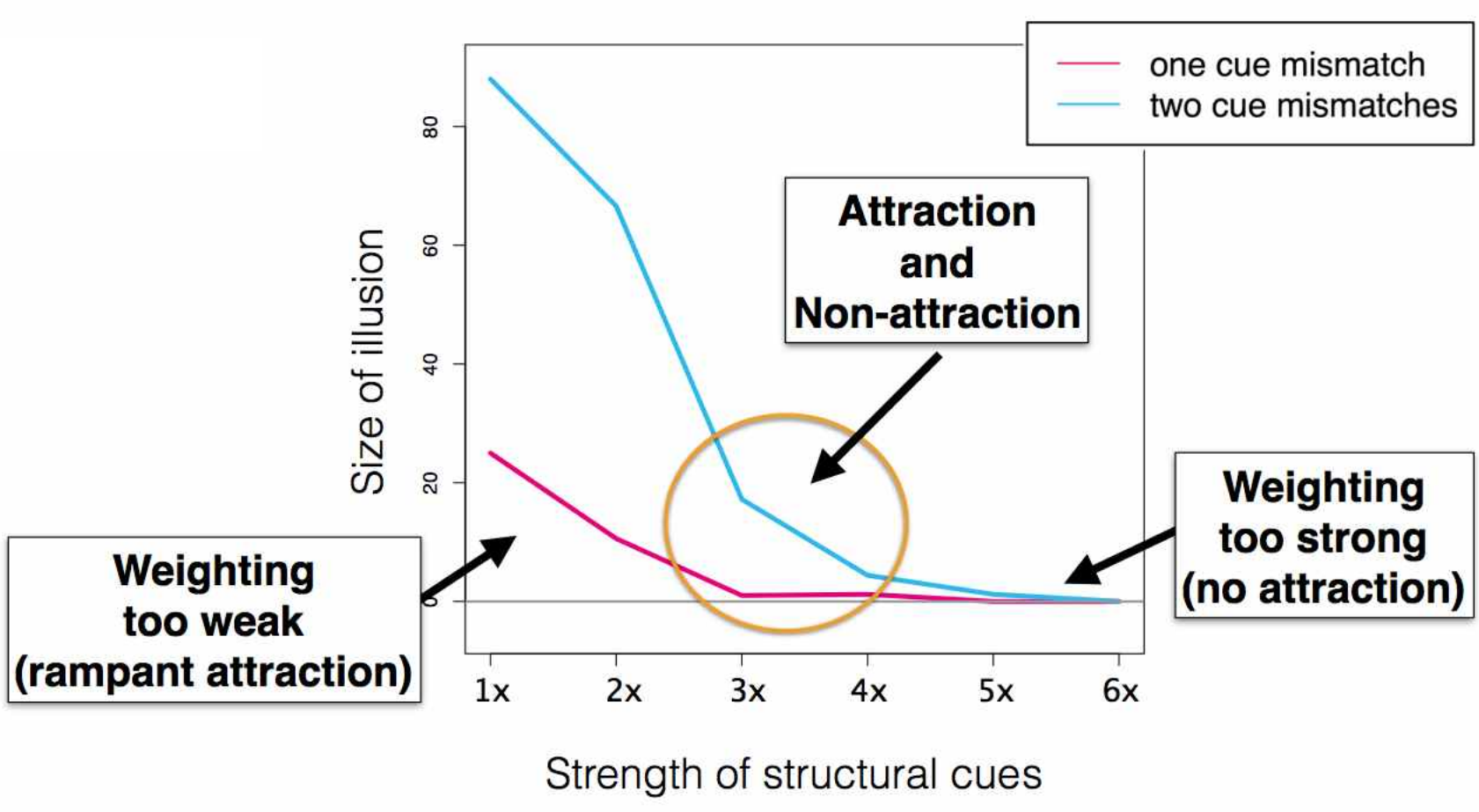
Revised Generalization

- **Previous**
  agreement is susceptible to illusions
  reflexive licensing is not
  → they deploy different memory access mechanisms

- **Revised**
  agreement and reflexives are both susceptible to illusions
  … but differentially susceptible
  → they deploy the same memory access mechanisms
  → but structural information is weighted more strongly for reflexives

- **Why do they differ?**
  Agreement → predictable; errors trigger retrieval, structure unreliable
  Reflexives → not predictable; retrieval happens anyway, structure reliable

ACT-R simulation data: structural vs. morphological cues



Structure is not prioritized for violated agreement predictions

Reflexives are susceptible to the same attraction effects as verbs . . . so why do verbs fail so much more easily?

Strong attraction effects even in a 1-feature mismatch context implies that structural cues are not privileged for computing subject-verb agreement.

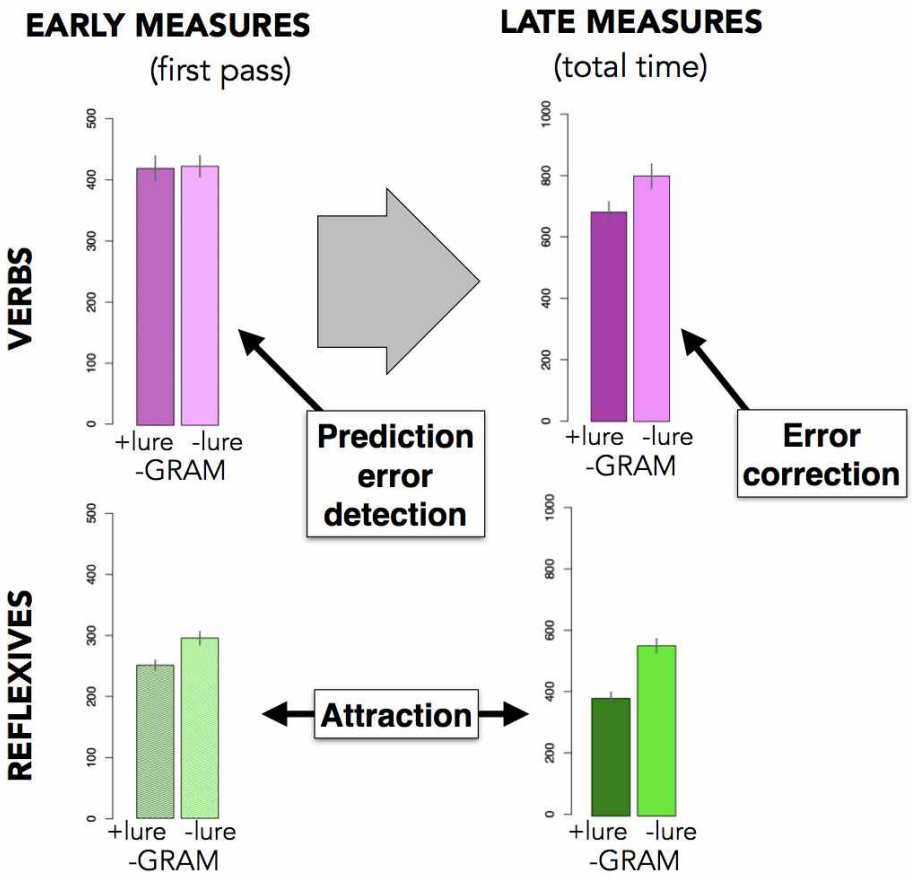
Proposal: Selective priority for structural cues reflects whether or not retrieval is trigged by prediction error

Agreement:
- predictable
- error triggers retrieval
- error reduces comprehenders' confidence in their structural encoding

Reflexives:
- not predictable
- retrieval happens anyway
- comprehenders are confident in their structural encoding

Different timing profiles for verbs and reflexives

(data from Experiment 2)



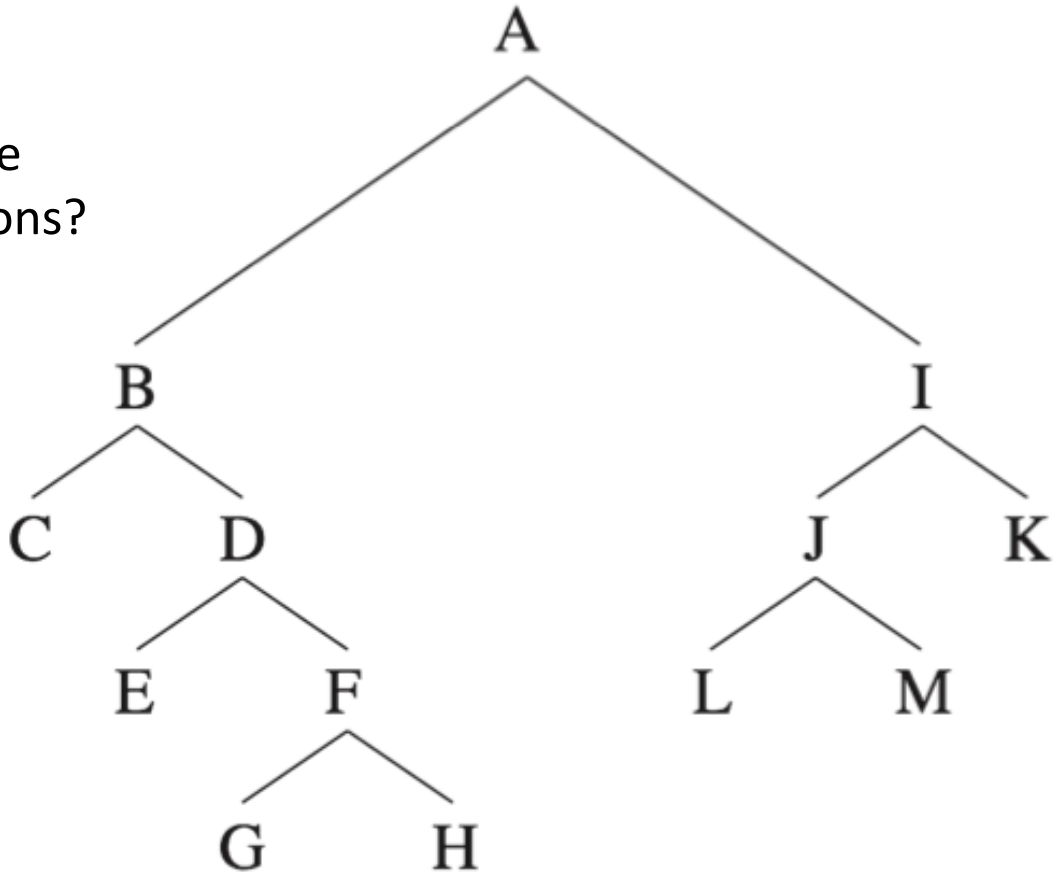
Dave Kush

Surprise #2

RELATIONAL ACCESS

Core Question

How do we encode and navigate structured mental representations?



Structures are easy to draw. But what kind of mental object are they?

Linguistic arguments: relational notions like c-command are pervasive.

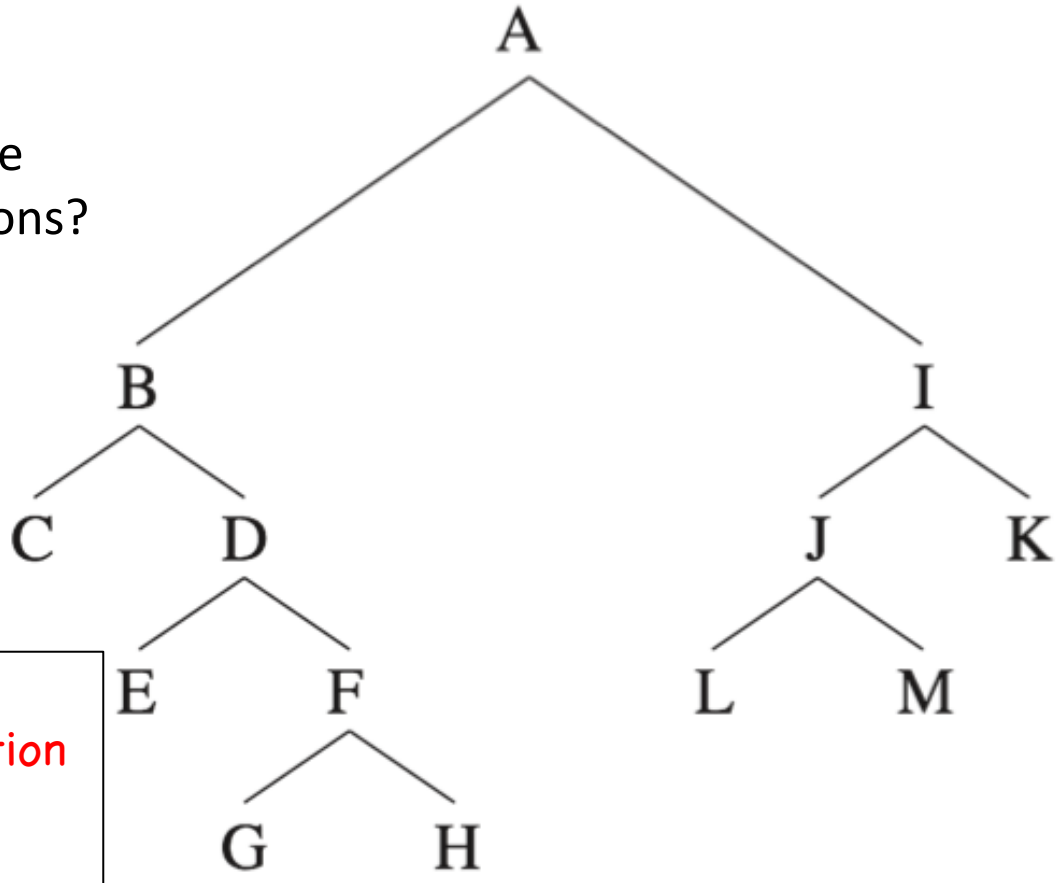
In some architectures it is trivial to capture relational notions like c-command.
In other architectures it is not trivial.

c-command constraints

- Bound variable anaphora
  - **No student$_i$** [that listened to this physics professor] thinks that **he$_i$** is a genius.
  - *The student [that listened to **no physics professor$_i$**] thinks that **he$_i$** is a genius.
- Many other relations : c-command is pervasive
  - Wh-dependencies
  - Reflexive binding
  - Negative polarity
  - Control
- Possible effects on-line
  - c-command serves “gating” function: only c-commanders accessed
  - c-command as non-exclusive constraint, on par with others (number, case, etc.)
  - c-command not directly implicated in retrieval at all

Core Question

How do we encode and navigate structured mental representations?



If we have
(i) Graph structure representation
(ii) Serial search mechanism
then c-command is cheap

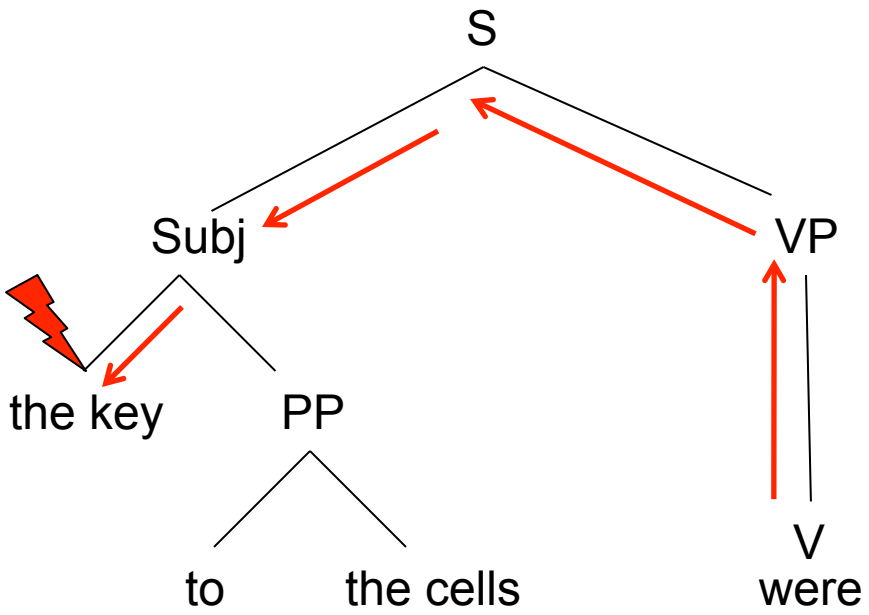
Structures are easy to draw. But what kind of mental object are they?

Linguistic arguments: relational notions like c-command are pervasive.

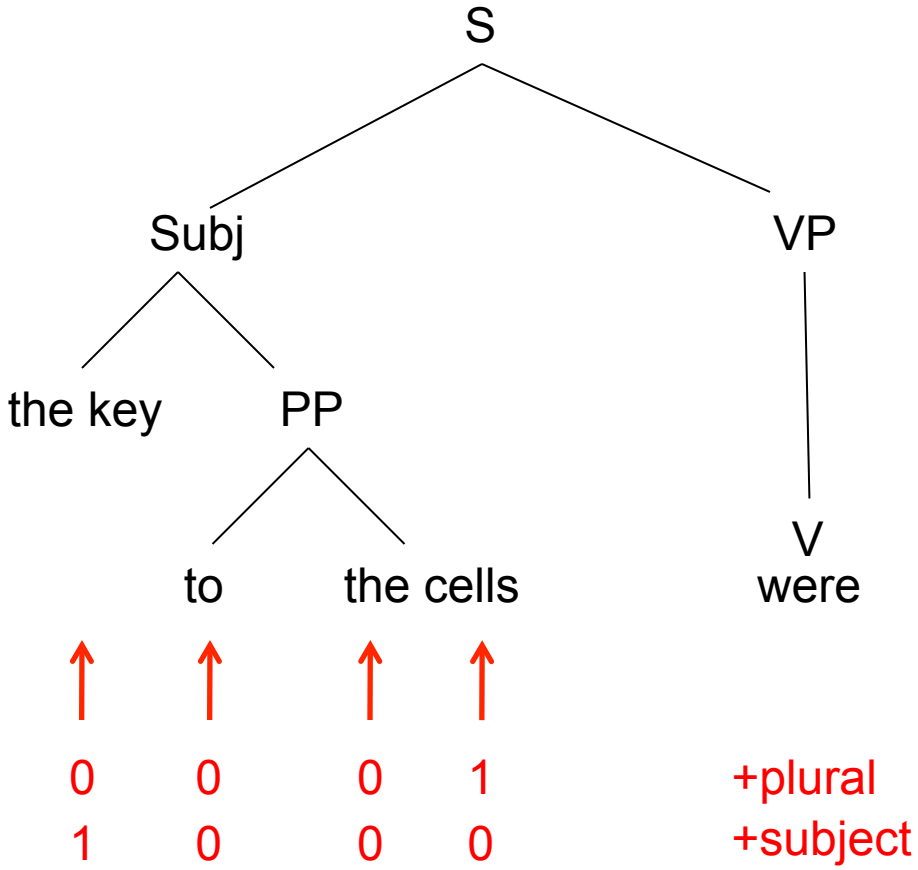
In some architectures it is trivial to capture relational notions like c-command.
In other architectures it is not trivial.

Two ways to search structures in memory

serial, structure-guided search



parallel, cue-guided (direct) access



structure-sensitive, avoids interference

slow, esp. for longer relations

susceptible to interference

fast, even for longer relations

McElree et al. 2003; Lewis et al. 2006; Wagers, Lau, & Phillips, 2009

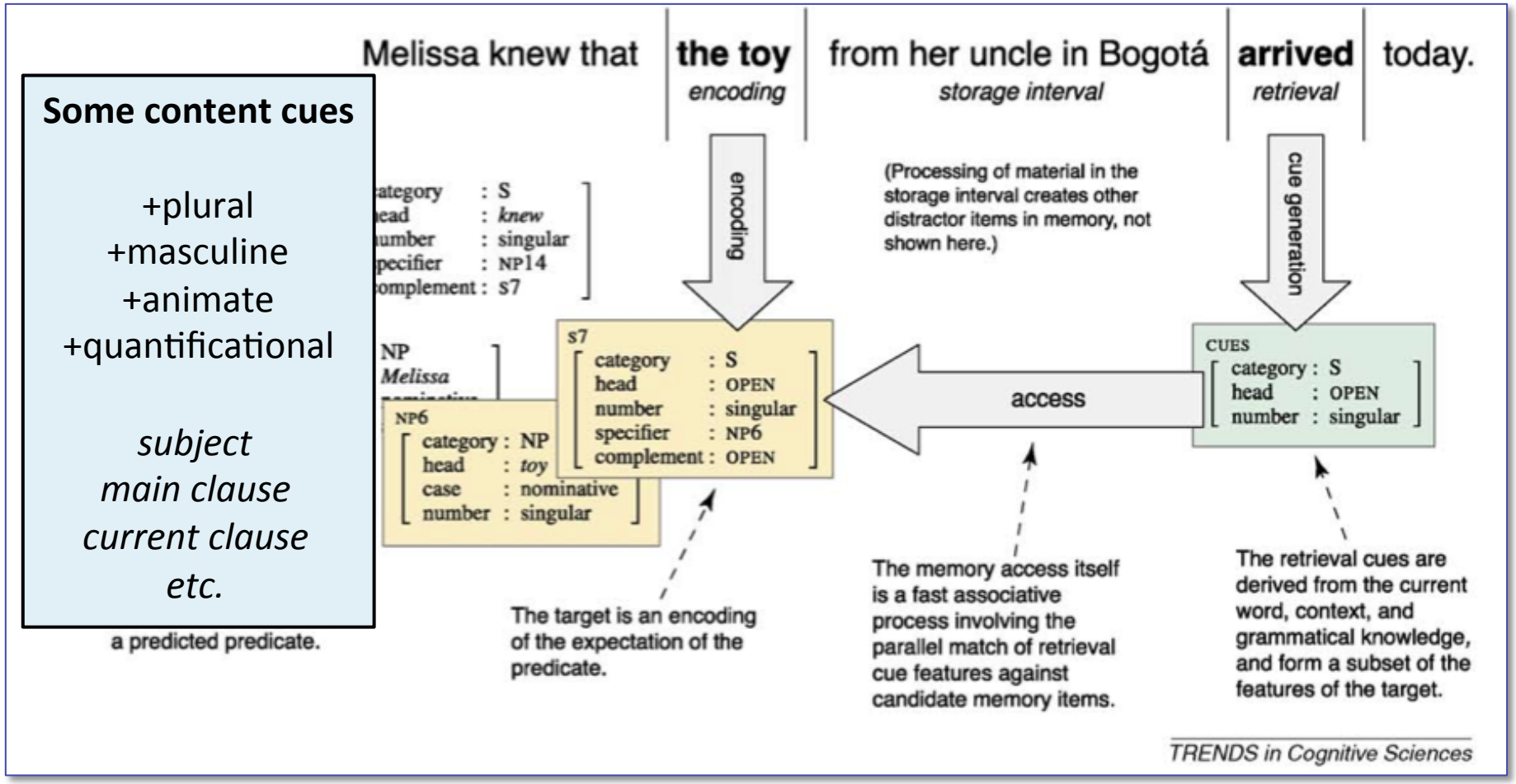
Memory-access Model

Key features

(i) Parallel-access → time-constant retrieval

(ii) Content-addressable → susceptible to partial-match interference

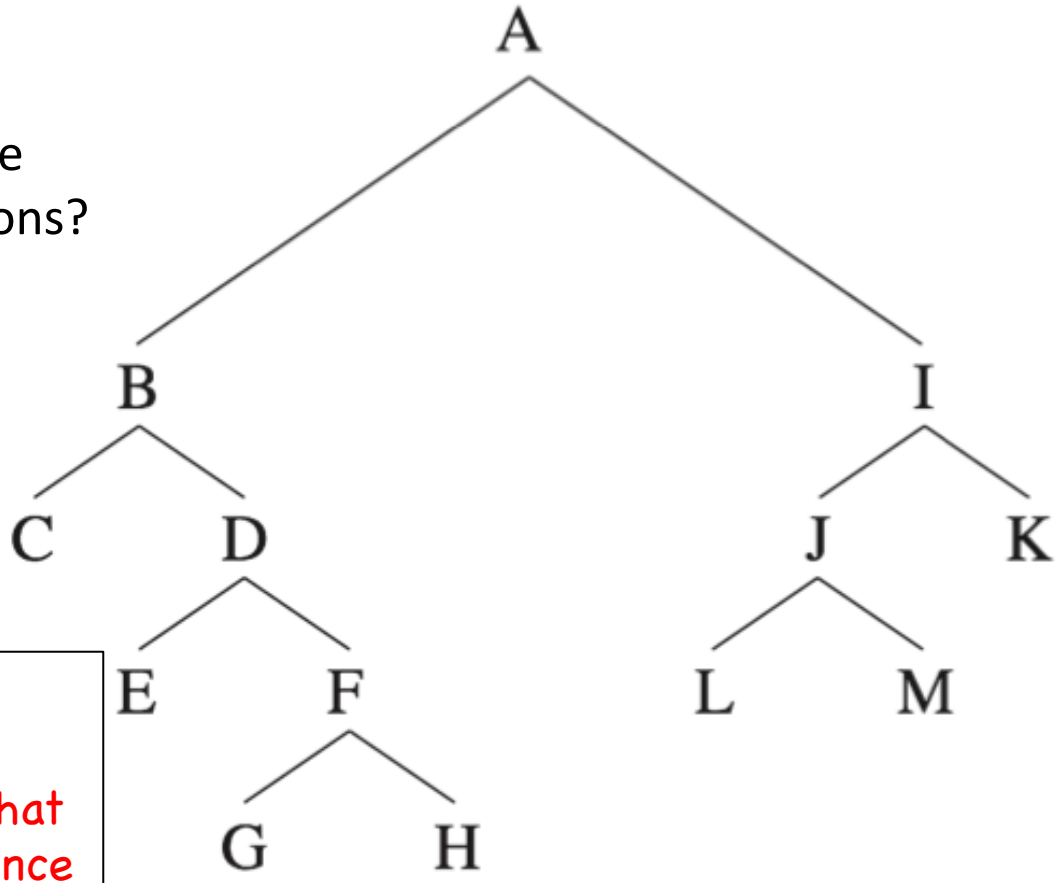
(iii) Limited-size buffers → ~ restricted focus of attention; const. shunting



(Lewis, Vasishth, & van Dyke 2006, *TICS*; cf. McElree 2006)

Core Question

How do we encode and navigate structured mental representations?



C-command is not content
"E c-commands H" is not an inherent property of node E, that exists independent of H's presence

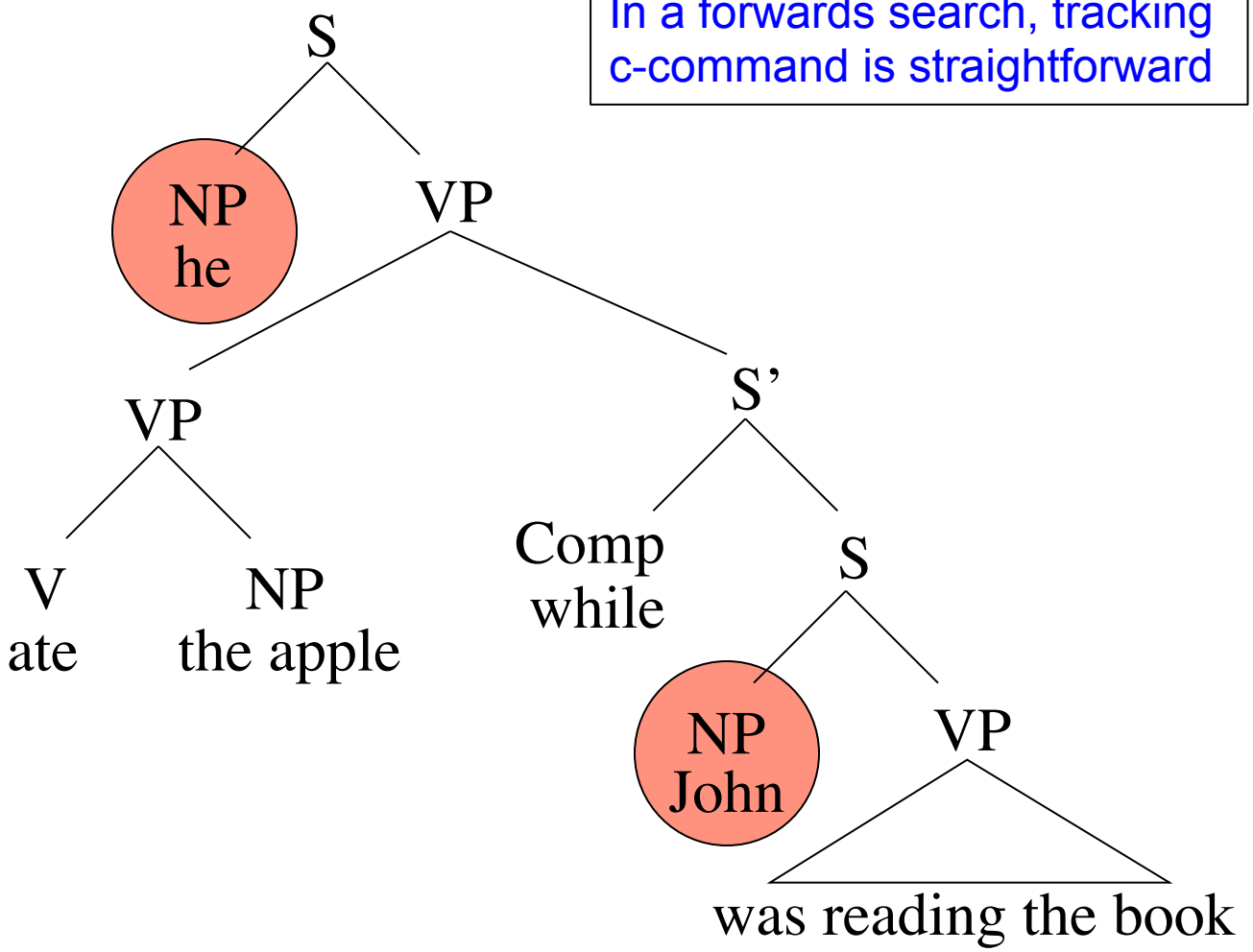
Structures are easy to draw. But what kind of mental object are they?

Linguistic arguments: relational notions like c-command are pervasive.

In some architectures it is trivial to capture relational notions like c-command.
In other architectures it is not trivial.

Principle C

In a forwards search, tracking c-command is straightforward



He ate the apple while John was reading the book

Binding vs. Coreference

No cyclist thought the spectators loved **him**.

The spectators [that **no cyclist** acknowledged]
thought the people loved **him**.

The cyclist thought the spectators loved **him**.

The spectators [that **the cyclist** acknowledged]
thought the people loved **him**.

(Kush, Lidz, & Phillips, 2015)

Binding vs. Coreference

2 x 2 Design: GenderMatch x Referentiality

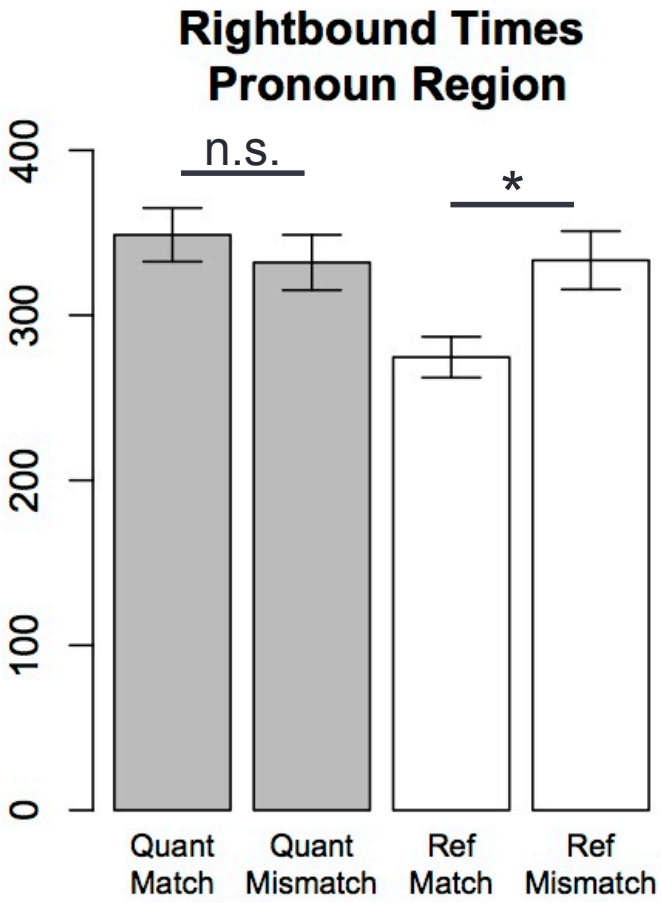
The troop leaders . . .

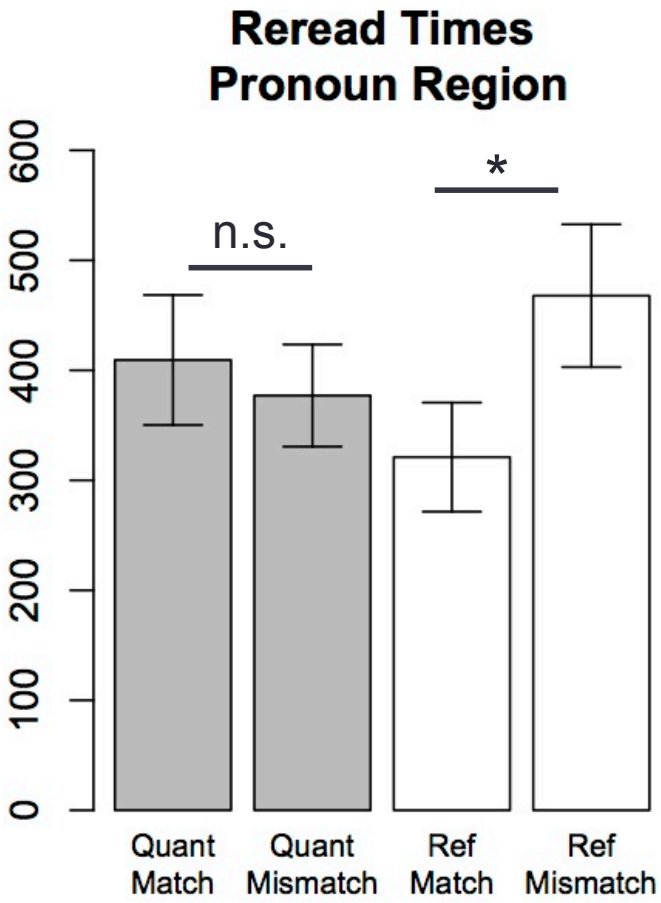
Match-Quant	[that **no girl scout** had respect for]
Mismatch-Quant	[that **no boy scout** had respect for]
Match-Ref	[that **the girl scout** had respect for]
Mismatch-Ref	[that **the boy scout** had respect for]

had scolded ***her*** after the incident …

(Kush, Lidz, & Phillips, 2015)

Binding vs. Coreference n = 30





C-command gates access to antecedents in retrieval
But what about Principle C effects in retrieval?

(Kush, Lidz, & Phillips, 2015)

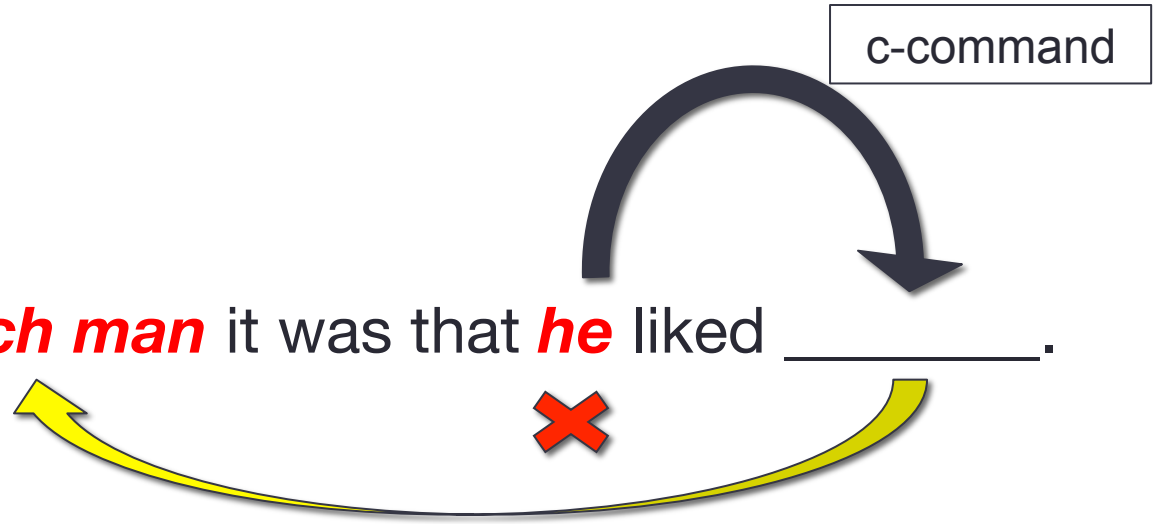
Strong Crossover

Mary knew ***which man*** __ said that ***he*** liked cheese.

Mary knew ***which man*** it was that ***he*** liked ___.

(Kush, Lidz, & Phillips, in prep)

Strong Crossover



*Mary knew ***which man*** it was that ***he*** liked _____.

(Kush, Lidz, & Phillips, in prep)

Strong Crossover and Retrieval

Jane asked …

NoCrossover Sentences

which janitor

which lunchlady

_____ had said that ***he*** had spoken with Donna ...

Jane asked …

Crossover Sentences

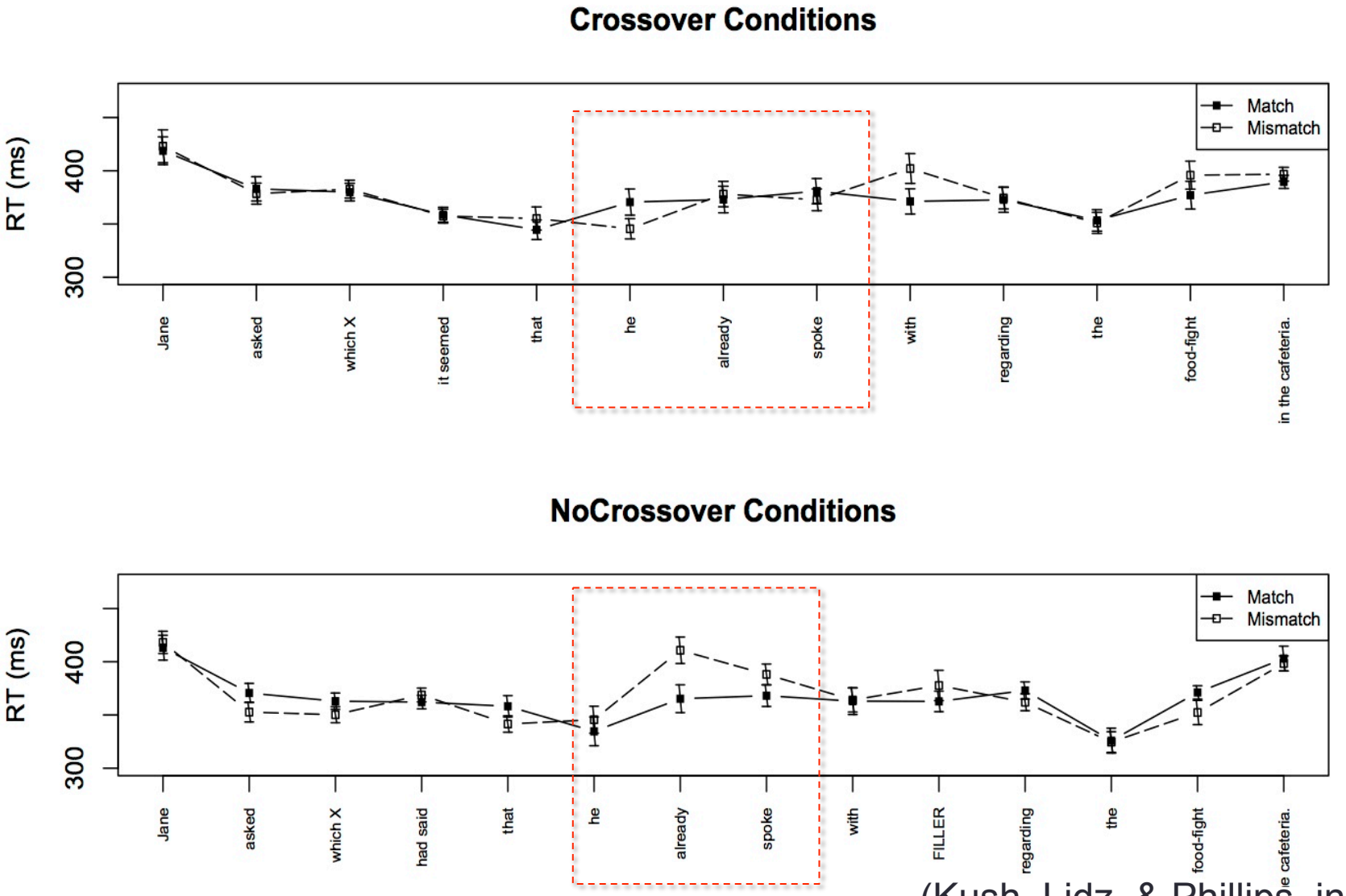
which janitor

which lunchlady

…it seemed that ***he*** had spoken with _____ ...

(Kush, Lidz, & Phillips, in prep)

Strong Crossover: SPR Times n=30

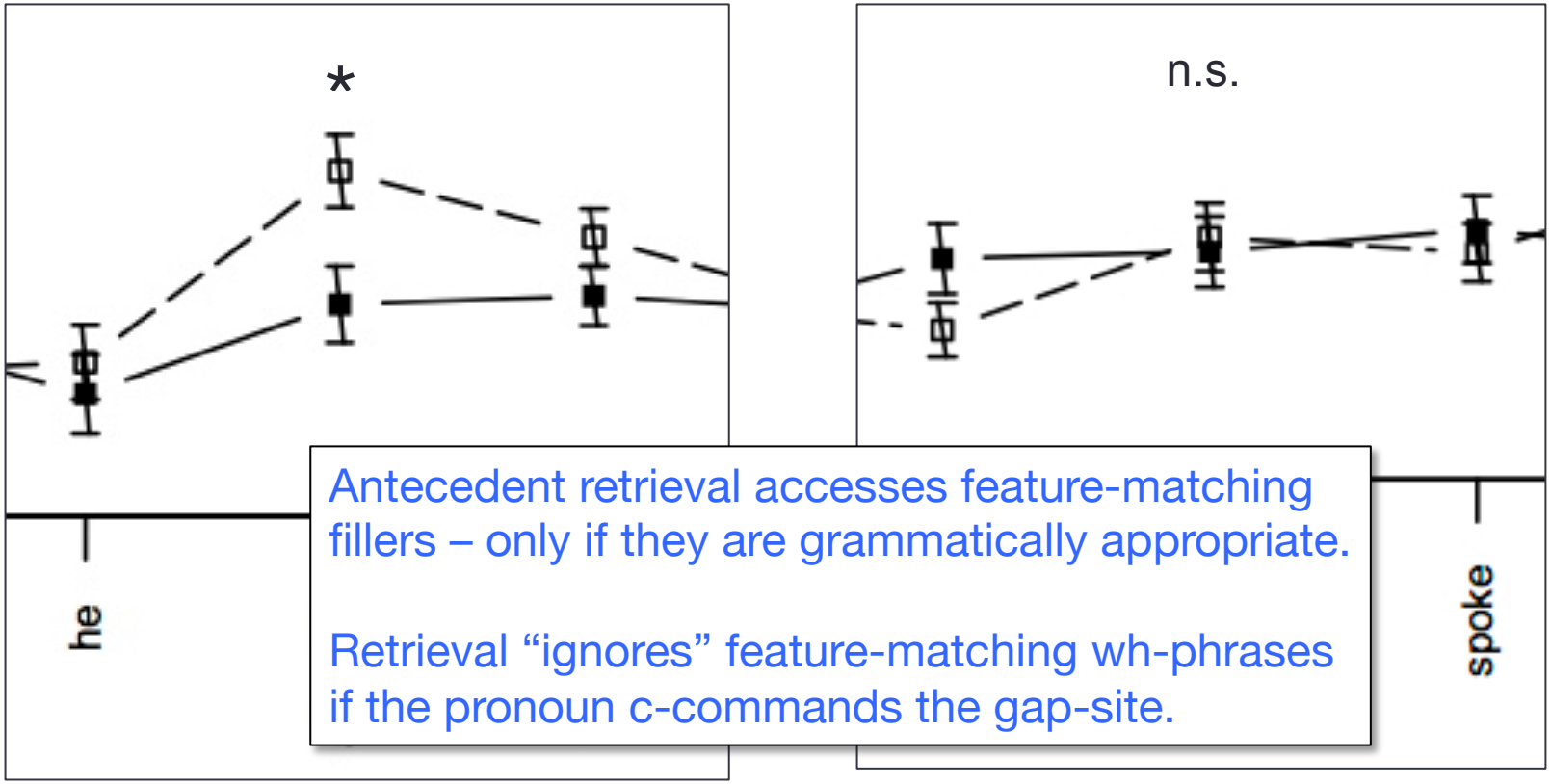


(Kush, Lidz, & Phillips, in prep)

Strong Crossover: Results

NoCrossover Conditions

Crossover Conditions



Antecedent retrieval accesses feature-matching fillers – only if they are grammatically appropriate.

Retrieval "ignores" feature-matching wh-phrases if the pronoun c-commands the gap-site.

(Kush, Lidz, & Phillips, in prep)

Strong Crossover and Retrieval

Jane asked …

NoCrossover Sentences

which janitor

which lunchlady

_____ had said that ***he*** had spoken with Donna ...

wh-gap before pronoun

Jane asked …

Crossover Sentences

which janitor

which lunchlady

wh-gap after pronoun

…it seemed that ***he*** had spoken with _____ ...

(Kush, Lidz, & Phillips, in prep)

Relational Access

- Real-time sensitivity to relational constraints
  - Retrieval can target items based on c-command (bound vbls)
  - Robustness of Principle C effects not due to forwards search
- What makes this possible?
  - Serial access mechanism? Easily captures c-command sensitivity.
    … but how to invoke only when QP antecedent is available?
  - Relational properties as content (via costly encoding function)?
  - Even cases that look like (backwards) retrieval involve forwards updating
    e.g., QP & whP induce tracking of c-command path (cf. Kush 2013 PhD)
- Why doesn't everything use this?
  - Varying usefulness/reliability of deploying structure-sensitive mechanisms
  - Feasibility of identifying elements likely to act as c-commanding licensors

Intervention

John found a picture of himself in the cabinet.

*John found Mary's picture of himself in the cabinet.

Zhangsan zhidao Lisi piping-le ziji.

Z. *knows* *L.* *criticized self*

*Zhangsan zhidao wo piping-le ziji.

Z. *knows* *I* *criticized self*

Nobody said that Bill would ever vote for Clinton.

*Nobody said that everybody would ever vote for Clinton.

What do you know that Sally likes __?

*What do you know who likes __?

Mandarin Long-Distance Reflexives

Lisi *nongshang-le* ***ziji***
Lisi harm-PERF self
“Lisi harmed herself”

Zhangsan *shuo* ***Lisi*** *nongshang-le* ***ziji***
Zhangsan says Lisi harm-PERF self
“Zhangsan says that Lisi harmed him / herself”

Antecedent for *ziji* must be
(i) a subject
(ii) animate
(iii) c-commanding

Speed-Accuracy Tradeoff (MR-SAT)

LD antecedent: **Coach Zhang** say [that report [..] underestimate ***ziji***]

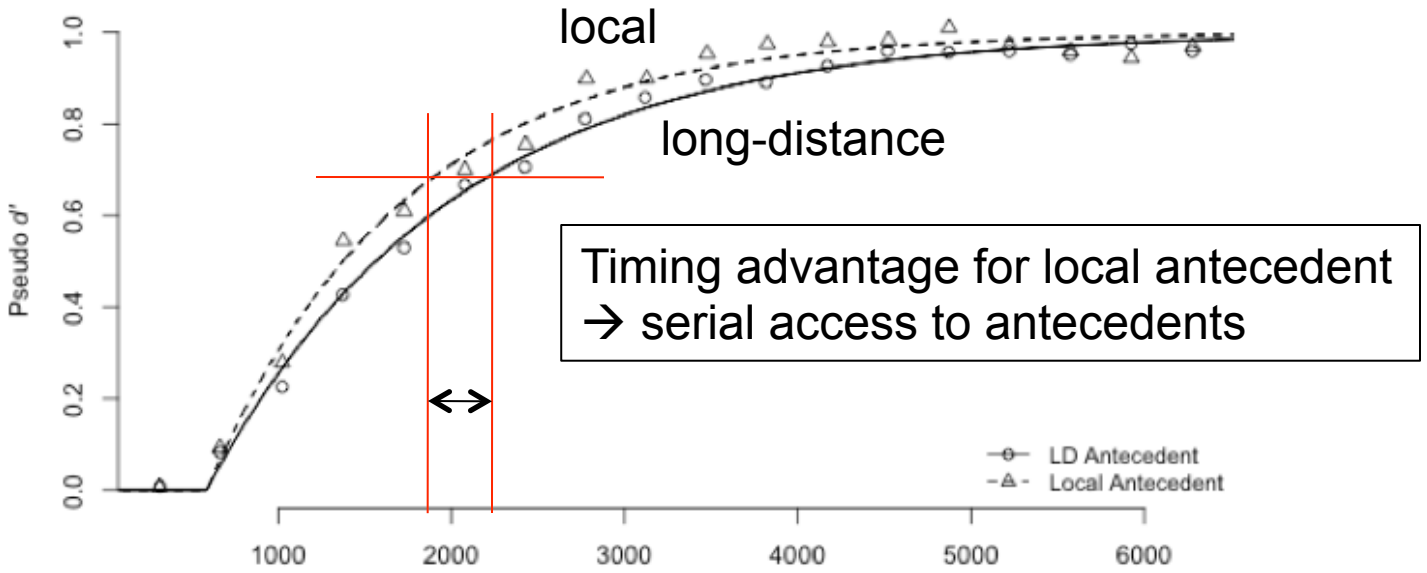
Local antecedent: Auto-biography say [**Coach Zhang** [..] underestimate ***ziji***]

No antecedent: Auto-biography say [that report [..] underestimate ***ziji***]

+ control conditions without *ziji*.

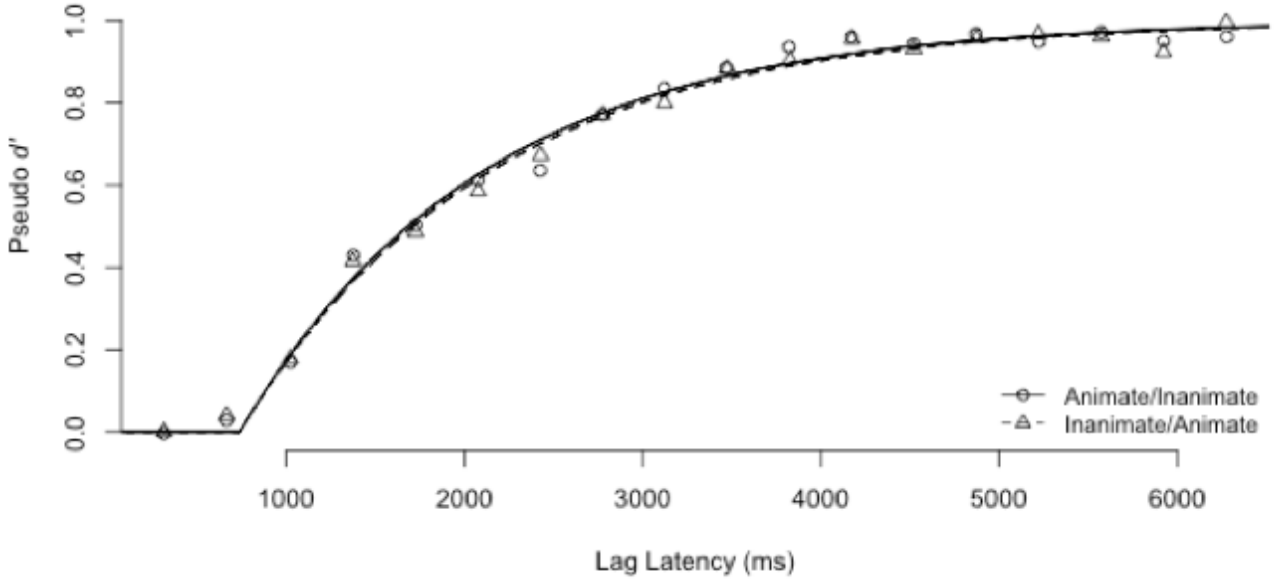
	Condition	Example
1	LD antecedent	张教练 表明 那篇报导［在 团队 未能 发挥 水准 的时候］低估了 自己 Coach Zhang say [that report [when team not perform well time] underestimate ziji] "Coach Zhang says that that report underestimated *self* when the team was doing poorly."
2	Local antecedent	回忆录 表明 张教练［在 团队 未能 发挥 水准 的时候］低估了 自己 Auto-biography say [coach Zhang [when team not perform well time] underestimate ziji] "The auto-biography says that coach Zhang underestimated *self* when the team was doing poorly."
3	No antecedent	*回忆录 表明 那篇报导［在 团队 未能 发挥 水准 的时候］低估了 自己 *Auto-biography say [that report [when team not perform well time] underestimate ziji] * "The auto-biography says that that report underestimated *self* when the team was doing poorly."

ziji conditions

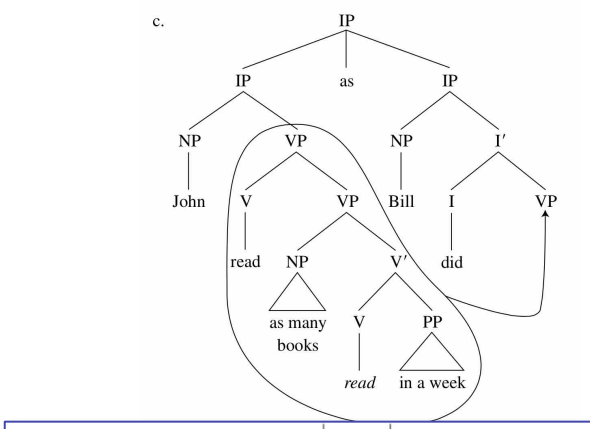


control conditions

(identical words, except for *ziji*)

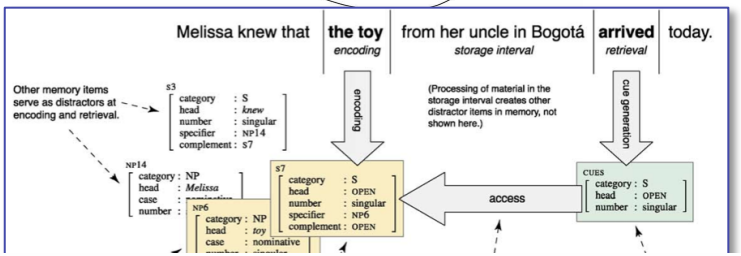


[common scaling of d' based on asymptote, for visualization of dynamics] (Dillon et al. 2014)



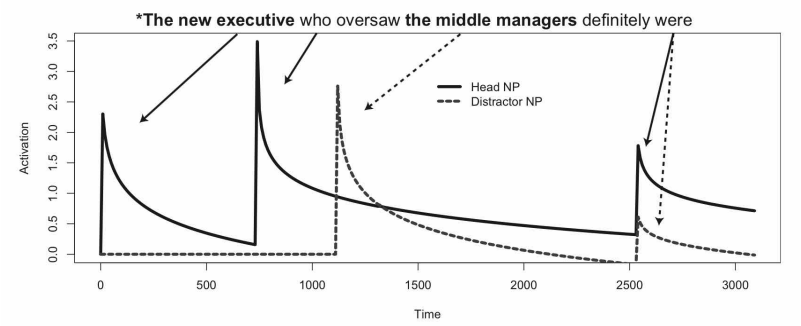
structural constraints

(left-right) order



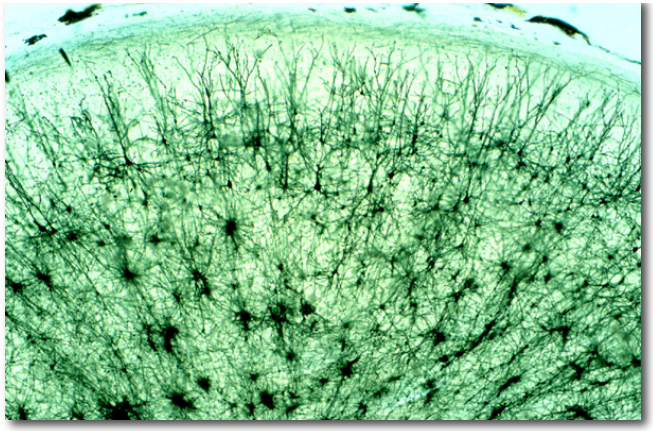
time

Structure & Memory



resource limitations

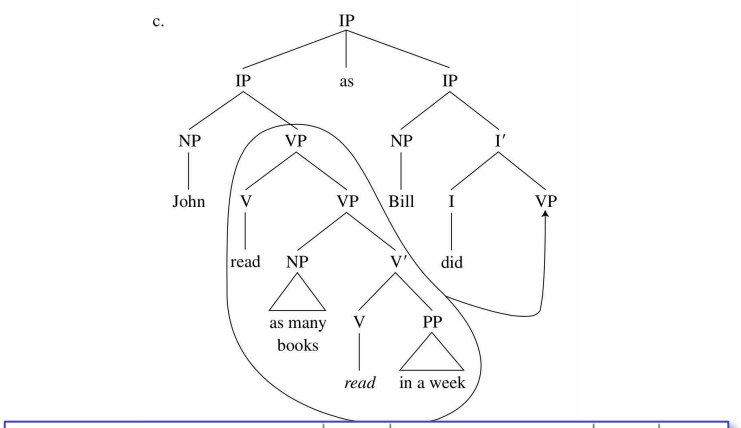
memory architecture and access



non-discrete units

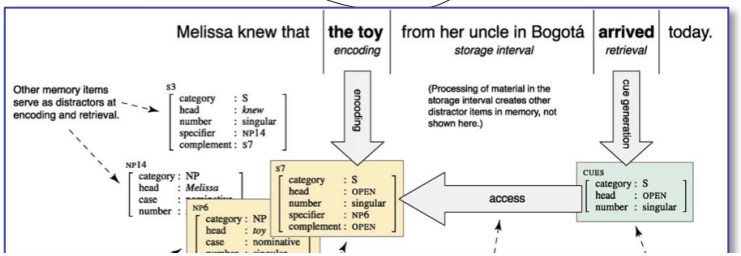
etc.

2. Implementation ≠ Reductionism



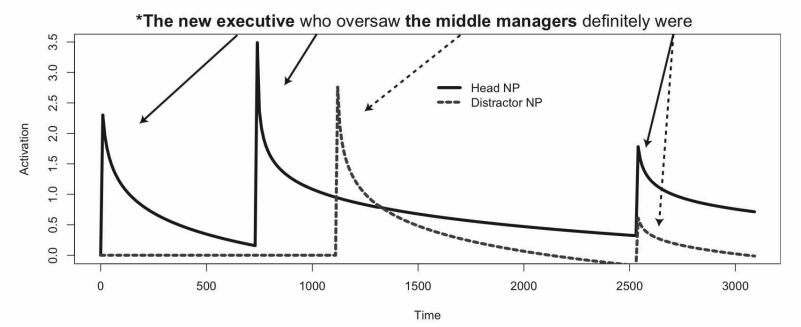
structural constraints

(left-right) order



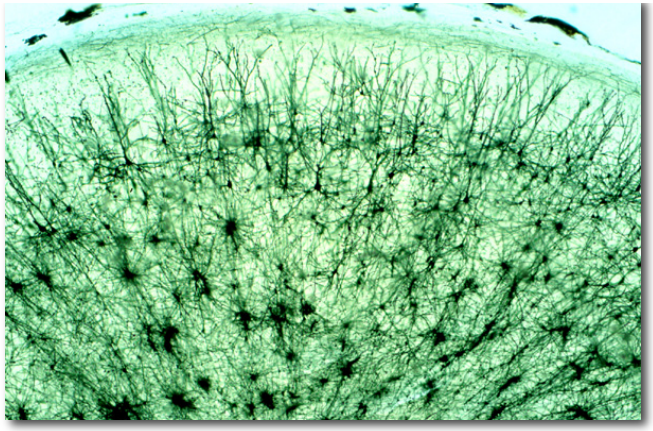
time

Blame Assignment



resource limitations

memory architecture and access



non-discrete units

etc.

Question Formation

Wallace suspects that everybody doubts that Gromit makes cracking toast.

Question Formation

Wallace suspects that everybody doubts that Gromit makes what.

Question Formation

What does Wallace suspect that everybody doubts that Gromit makes

Question Formation

relative clause

What does Wallace suspect anybody who doubts that Gromit makes

A relative clause is an ‘island’ – a wh-word cannot escape from it.

Island Constraints

5 a. *What did Wallace meet a woman [$_{\text{rel. cl.}}$ that hates ___]?
b. *What did John read the report [that Craig won ___]?
c. * Who did Robyn believe [Simon's news about ___]?
d. *What did Sue wonder [whether Joe wrote ___]?
e. *What does Helen know [who saw ___]?
f. *What did [the fact that Ellen remembered ___] surprise her children?
g. *Who did Susan watch TV [while talking to ___ on the phone]?
h. *What did [the Senate approve ___] and [the House reject the bill]?
i. *Why did they remember that the corrupt CEO had been acquitted ___?
j. *Why did they say that nobody left ___?

A common inference

- **Grammatical generalization:** wh-dependencies are local
- **Parsing generalization(s):** local wh-dependencies are easier/preferred
- ***ERGO…*** perhaps the grammatical generalization derives from the parsing generalization
  - **Variant I:** locality constraints are nevertheless grammaticized
    Fodor 1978, Berwick and Weinberg 1984, Hawkins 1999
  - **Variant II:** locality constraints in grammar are epiphenomenal
    Kluender and Kutas 1993, 1998, 2004, Sag et al 2005; Hofmeister & Sag 2009, Maratsos and Kowalsky 2005


Fodor


Weinberg


Berwick


Hawkins


Kluender


Hofmeister


Sag


Maratsos

One way of testing time course of question interpretation

!!

How many students did the school enlarge the classroom for __?

Question Formation

relative clause

What does Wallace suspect anybody who doubts that Gromit makes

A relative clause is an 'island' – a wh-word cannot escape from it.

Does the constraint impact real-time comprehension of questions? Yes!

Does this tell us what is the source of island effects? No!

Stowe 1986; Traxler & Pickering 1996; Phillips 2006; Wagers & Phillips, 2009; Yoshida, Aoshima, & Phillips 2004

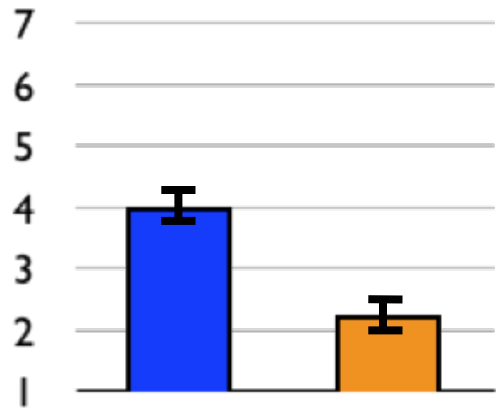
The logic of resource limitation theories

processing load 1 + **processing load 2** + **limited resources** = **an even greater processing disruption**

processing load 1: Longer wh-dependencies are harder to process than shorter dependencies (LENGTH)

John thought that you knew who the president would pardon ___ ?

John knew who you thought that the president would pardon ___ ?



There are many potential reasons for this: memory encoding, or maintenance, or retrieval...

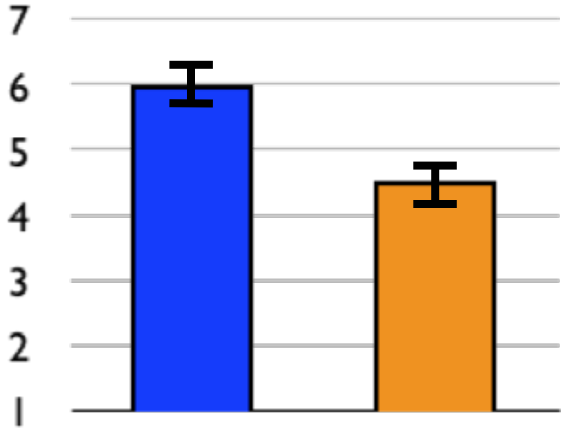
Crucially, longer dependencies lead to lower acceptability

The logic of resource limitation theories

processing load 1 + **processing load 2** + **limited resources** = **an even greater processing disruption**

processing load 2: Island structures are harder to process than non-island structures (STRUCTURE)

Who ___ thinks [CP that John bought a car]?
Who ___ wonders [CP whether John bought a car]?



(Kluender and Kutas 1993b, Sprouse 2007)

There are many potential reasons for this: referential processing, syntactic complexity, semantic complexity...

Crucially, island structures without island violations lead to lower acceptability

Argument #1: Individual differences

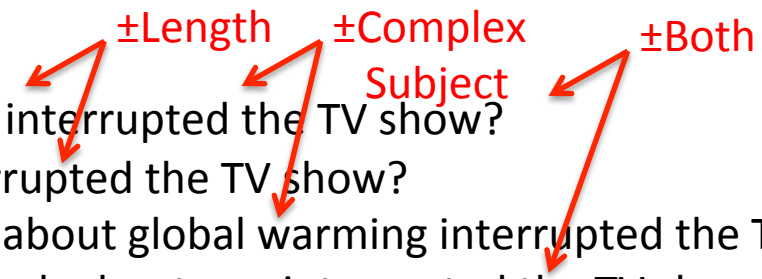
- Theoretical claim: island violations are not ungrammatical, just difficult
  (e.g., Kluender & Kutas 1993; Hofmeister & Sag 2010)

- **if** island effects reflect processing capacity overload,
  then severity of island effects should covary with individual capacity

- Island: interaction of (i) long extraction, (ii) island-inducing structures

Subject island

a. Who ___ thinks the speech interrupted the TV show?
b. Who do you think ___ interrupted the TV show?
c. Who ___ thinks the speech about global warming interrupted the TV show?
d. *Who do you think the speech about ___ interrupted the TV show?



Sprouse, Wagers, & Phillips 2012, *Language*

(6) Complex NP islands

a.	Who __ claimed that Lily forgot the necklace?	MATRIX | NON-ISLAND
b.	What did the teacher claim that Lily forgot __?	EMBEDDED | NON-ISLAND
c.	Who __ made the claim that Lily forgot the necklace?	MATRIX | ISLAND
d.	*What did the teacher make the claim that Lily forgot __?	EMBEDDED | ISLAND

(7) Subject islands

a.	Who __ thinks the necklace is expensive?	MATRIX | NON-ISLAND
b.	What does Jack think __ is expensive?	EMBEDDED | NON-ISLAND
c.	Who __ thinks the necklace for Lily is expensive?	MATRIX | ISLAND
d.	*Who does Jack think the necklace for __ is expensive?	EMBEDDED | ISLAND

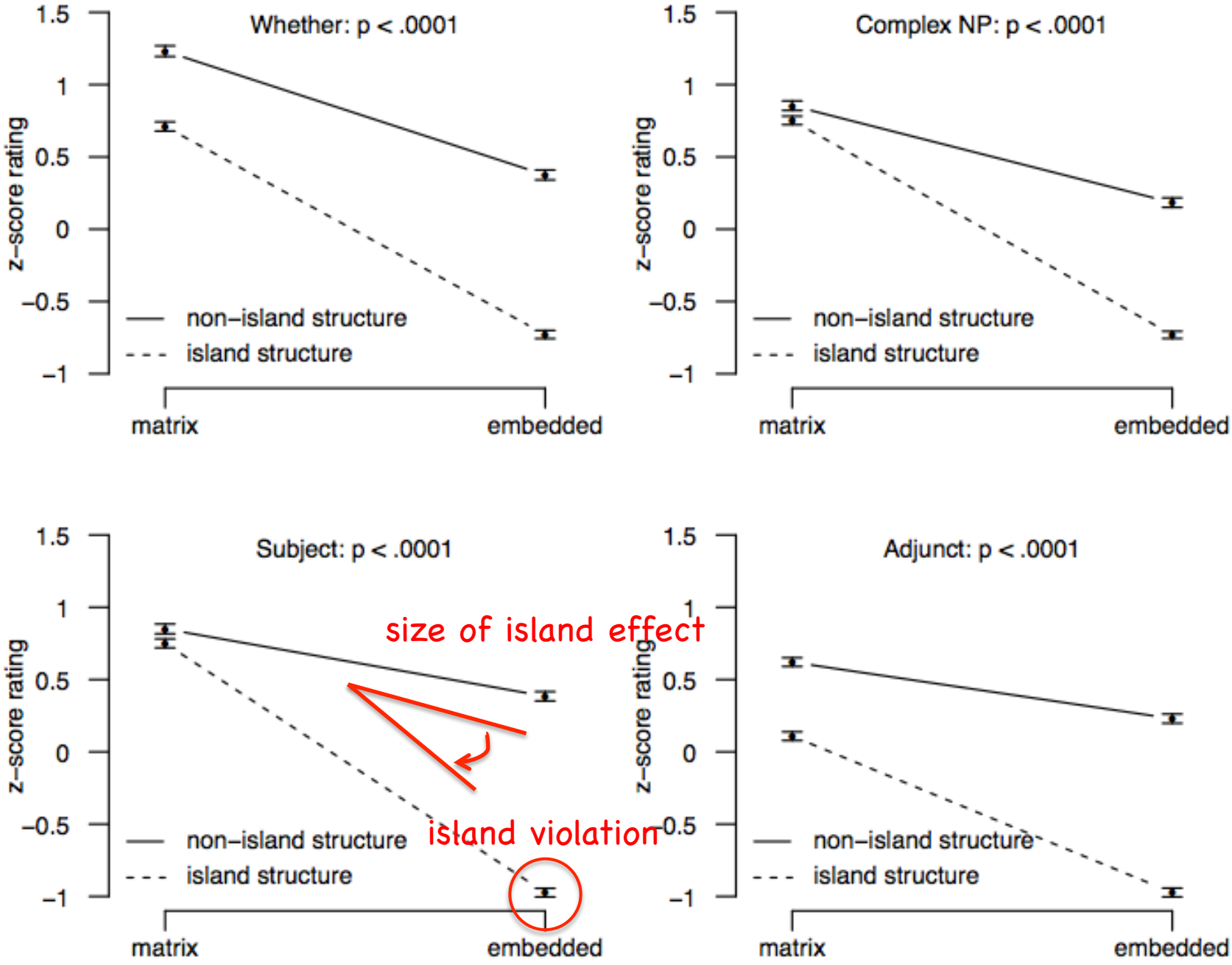
(8) Whether islands

a.	Who __ thinks that Jack stole the necklace?	MATRIX | NON-ISLAND
b.	What does the teacher think that Jack stole __ ?	EMBEDDED | NON-ISLAND
c.	Who __ wonders whether Jack stole the necklace?	MATRIX | ISLAND
d.	*What does the teacher wonder whether Jack stole __ ?	EMBEDDED | ISLAND

(9) Adjunct islands

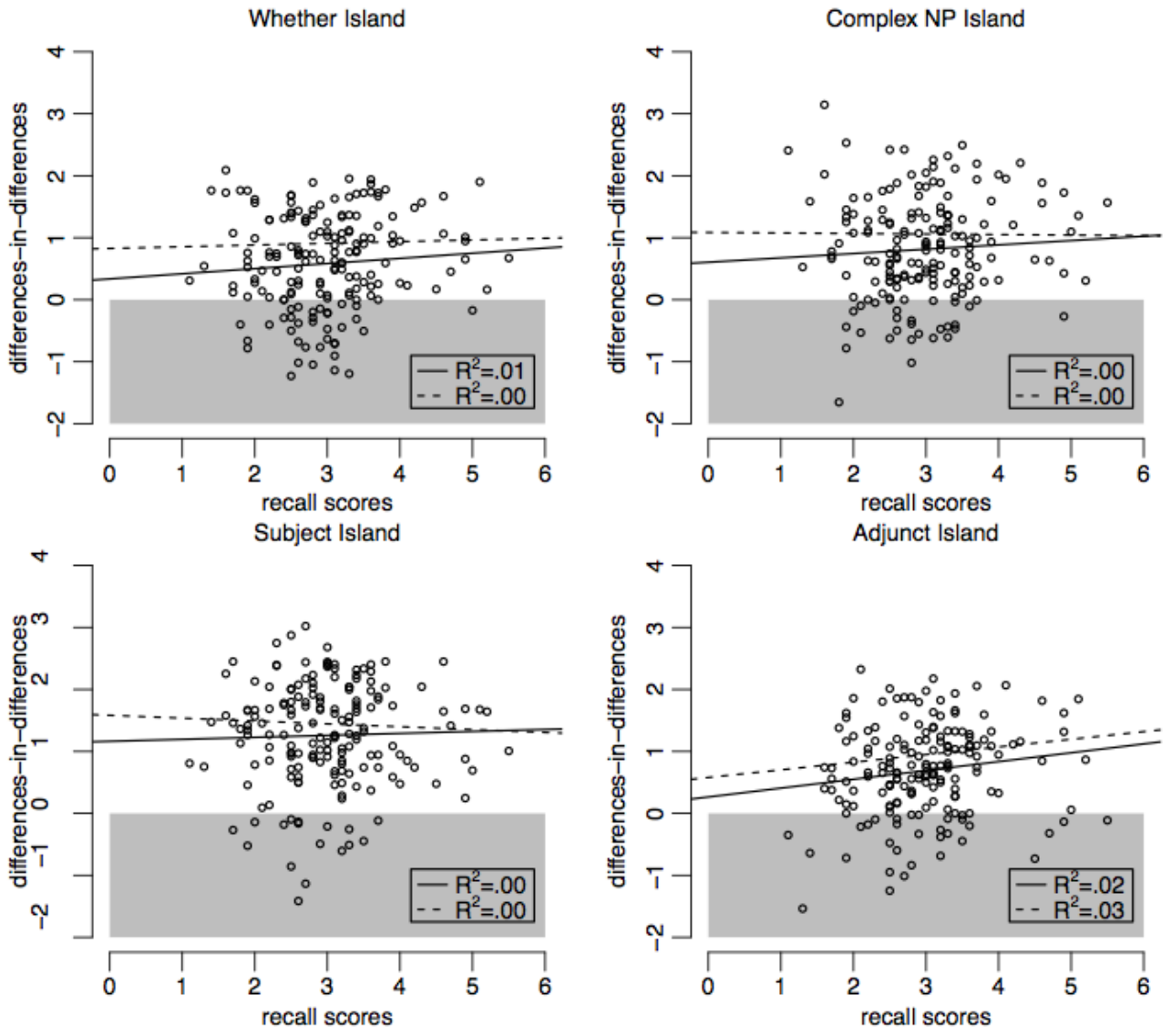
a.	Who __ thinks that Lily forgot the necklace?	MATRIX | NON-ISLAND
b.	What does the teacher think that Lily forgot __ ?	EMBEDDED | NON-ISLAND
c.	Who __ worries if Lily forgot the necklace?	MATRIX | ISLAND
d.	*What does the teacher worry if Lily forgot __ ?	EMBEDDED | ISLAND

Sprouse, Wagers, & Phillips 2012, *Language*



Sprouse, Wagers, & Phillips 2012, *Language*

4 island types, 2 memory tasks (serial recall, n-back), n = 315
island severity vs. memory capacity
capacity differences account for 0% - 3% of variance



Sprouse, Wagers, & Phillips 2012, *Language*

Argument #2

Yes, you can complete those dependencies on line
… when the grammar allows it

Phillips 2006, *Language*

Parasitic Gaps

the proposal to expand the school ultimately overburdened the teachers.

Phillips 2006, *Language*

Parasitic Gaps

the proposal to expand the school ultimately overburdened the teachers.

Phillips 2006, *Language*

Parasitic Gaps

which people did the proposal to expand the school ultimately overburden

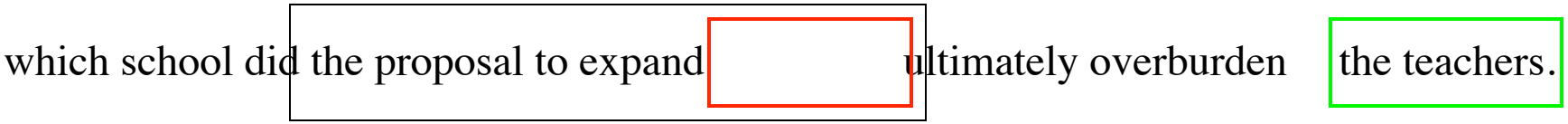
Phillips 2006, *Language*

Parasitic Gaps

which school did the proposal to expand ultimately overburden the teachers.

Phillips 2006, *Language*

Parasitic Gaps



Generalization (*Subject Island Constraint)*
No long-distance dependencies across subject boundaries

Phillips 2006, *Language*

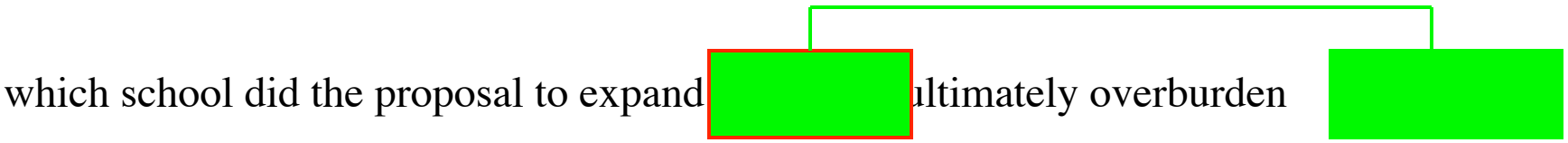
Parasitic Gaps

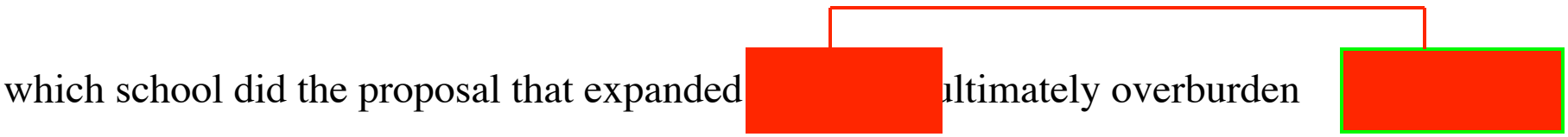
which school did the proposal to expand [ ] ultimately overburden [ ] .

Generalization (informal)
Violations can be rescued by subsequent well-formed gaps.

Phillips 2006, *Language*

Parasitic Gaps

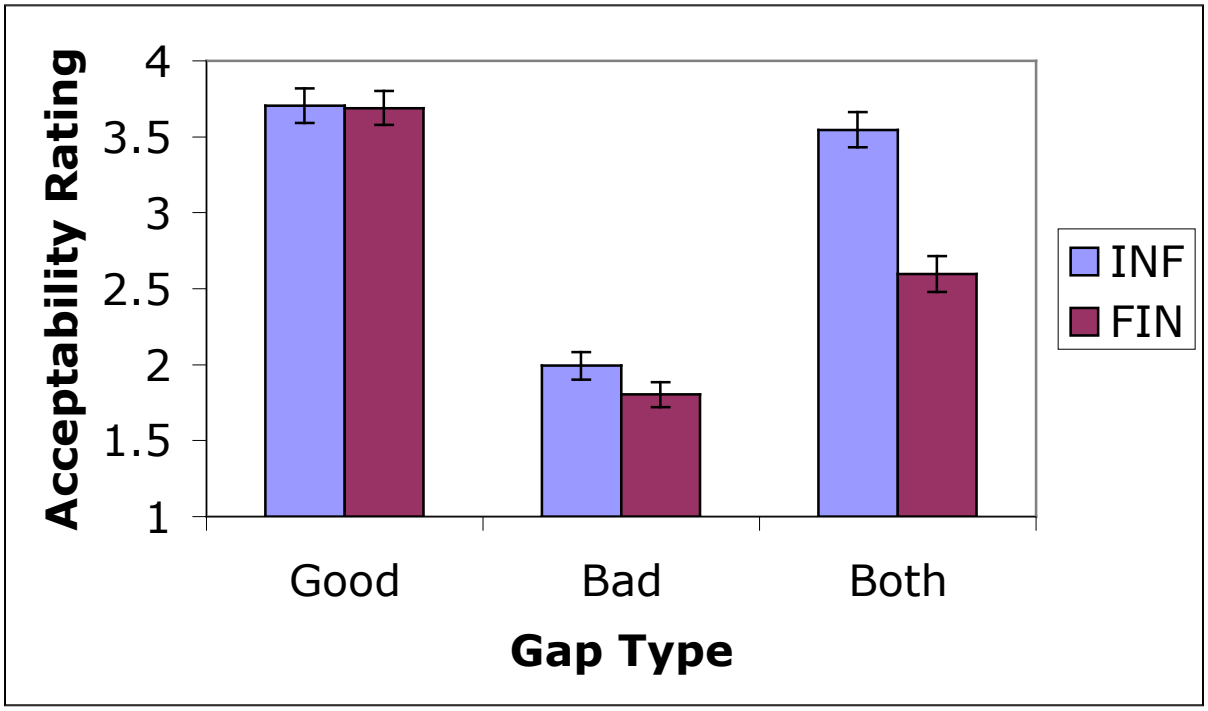




Updated Generalization (informal)
A *subclass* of violations can be rescued by subsequent gaps.

Phillips 2006, *Language*

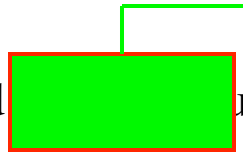
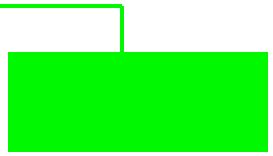
Acceptability Ratings



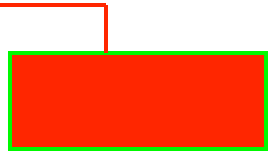
Phillips 2006, *Language*

Parasitic Gaps

plausible at ‘expand’
plausible at ‘overburden’

which school did the proposal to expand  ultimately overburden 
which students…

implausible at ‘expand’
plausible at ‘overburden’

which school did the proposal that expanded  ultimately overburden 
which students…

Phillips 2006, *Language*

Implications - ‘Parsing Accounts’

Any ‘processing based’ account of why this is bad…

which school did the proposal to expand ____ ultimately overburden the teachers.

…will fail to explain why the first gap can be created here…

which school did the proposal to expand ____ ultimately overburden ____ .

Phillips 2006, *Language*

(a) *Genitive/gender match*
His managers revealed that the studio that notified Jeffrey Stewart about the new film selected a novel for the script, but Annie did not seem to be interested in this information.

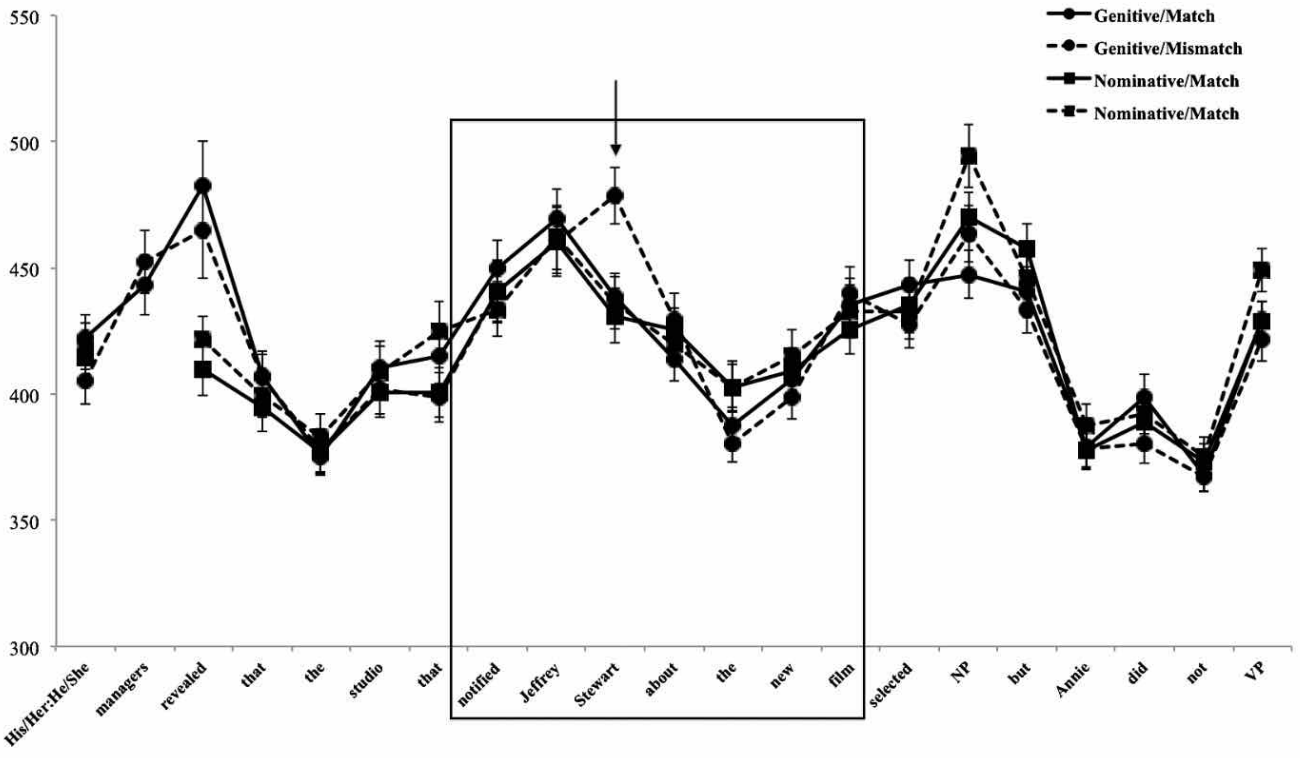
(b) *Genitive/gender mismatch*
Her managers revealed that the studio that notified Jeffrey Stewart about the new film selected a novel for the script, but Annie did not seem to be interested in this information.

(c) *Nominative/gender match*
He revealed that the studio that notified Jeffrey Stewart about the new film selected a novel for the script, but Andy did not know which one.

(d) *Nominative/gender mismatch*
She revealed that the studio that notified Jeffrey Stewart about the new film selected a novel for the script, but Annie did not know which one.

Argument #3: Other Dependencies

(Yoshida, Kazanina, Pablos, & Sturt, 2013)



Argument #4: Cross-language Variation

- English

  *The man [who$_i$ [the suit [$_{RC}$ that ___$_i$ is wearing]] is dirty] arrived late.

- Japanese

  kiteiru yoohuku-ga yogoreteiru sinsi
  is.wearing suit.nom dirty.is gentleman

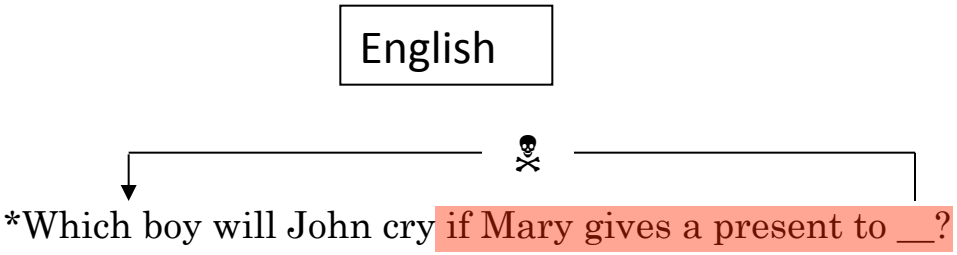
- Major Subject Construction (Japanese, Korean, Chinese)

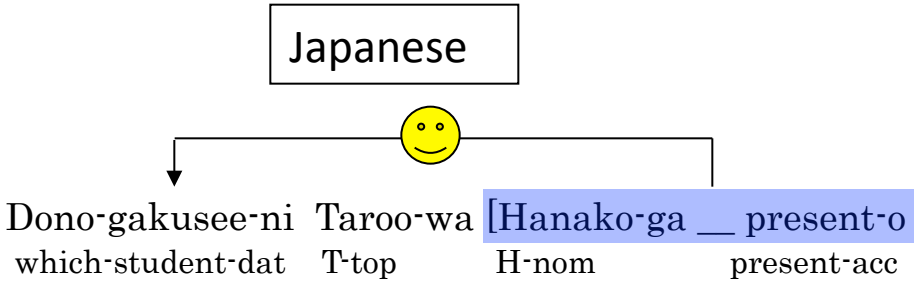
  [$_{IP}$ sono sinsi$_i$-ga [$_{NP}$ [$_{CP}$ pro$_i$ ___$_j$ kiteiru] [yoohuku$_j$]]-ga yogoreteiru]
  that gentleman-NOM pro wearing-is suit-NOM dirty-is
  'That gentleman is such that the suit that he is wearing is dirty.'

- [$_{CP}$ Op$_i$ [$_{IP}$ ___$_i$ [$_{NP}$ [$_{CP}$ pro$_i$ ___$_j$ kiteiru] yoohuku$_j$]-ga yogoreteiru] [sinsi$_i$]]
  Op pro wearing-is suit-NOM dirty-is gentleman
  'The gentleman who the suit that he is wearing is dirty.'

Argument #4: Cross-language Variation

Conditional Adjunct Clauses are islands in English, but not in Japanese.





Yoshida 2006, *thesis*

Active wh-gap creation and (un)successful revision in children

Akira Omaki
Johns Hopkins U

Omaki, Davidson White, Goro, Lidz, & Phillips, 2014

wh-attachment ambiguity

Where did Emily tell someone that she hurt herself?

1. Main Clause interpretation: where 'telling' happened
2. Embedded clause interpretation: where 'hurting' happened

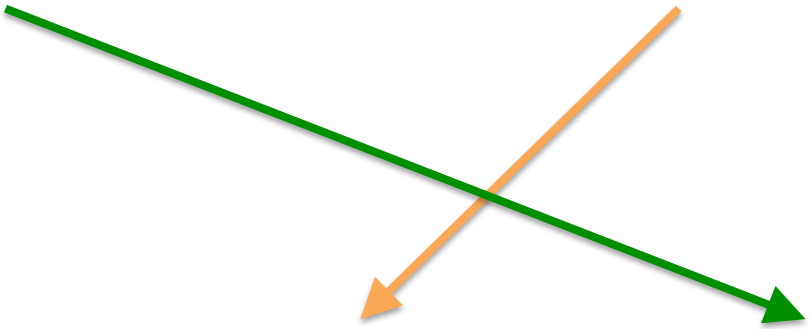
If the child parser has the following properties:

1. **Active gap creation** (e.g., Garnsey et al 1989; Stowe 1986; Yoshida & Dickey 2008)
2. **Reanalysis difficulty** (e.g., Choi & Trueswell 2010; Trueswell et al. 1999)

→ Main clause interpretation would be preferred

English vs. Japanese

Where did Emily tell someone that she hurt herself?

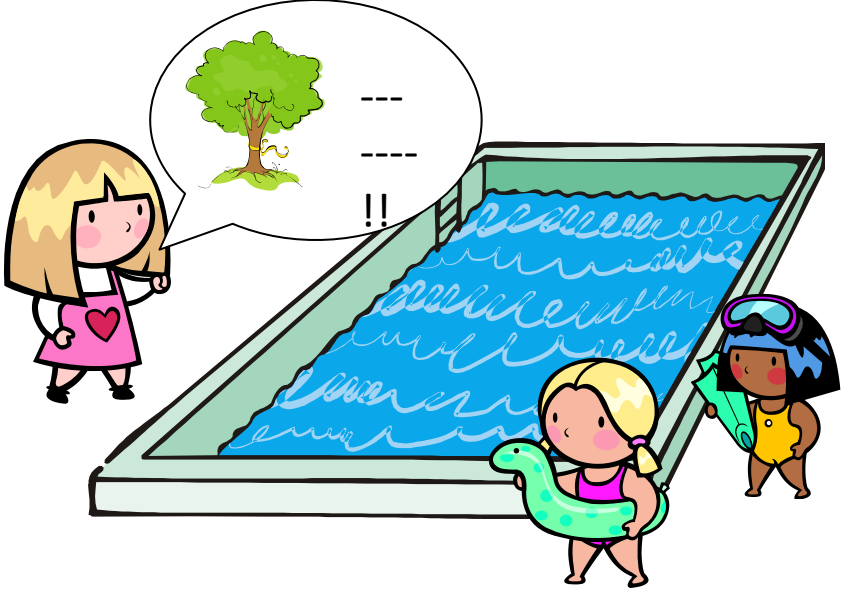


Doko-de	Emily-wa	[pro	asi-o	kegasita	to]	itta-no ?
Where-at	Emily-top	[pro	leg-acc	hurt	comp]	told-Q ?

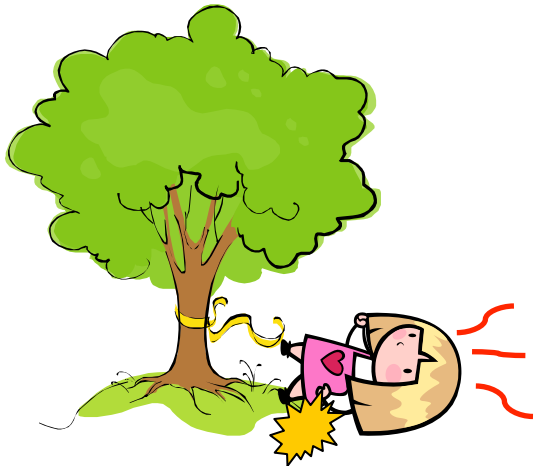
Adult reading-time evidence for **first verb association**

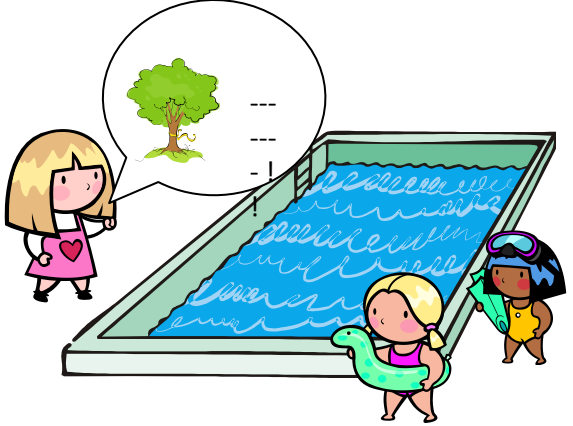
(Aoshima et al., 2004; Nakano et al., 2002)


LIBRARY



!!



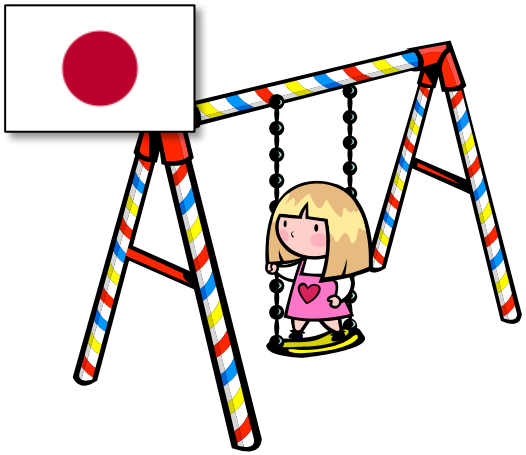


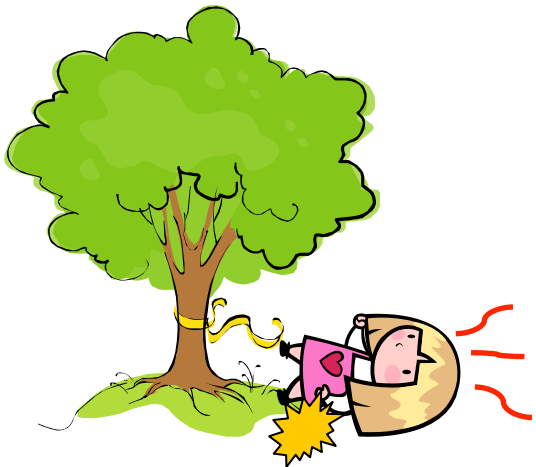
Question:

Where did Emily tell someone that she hurt herself?

POOL

TREE

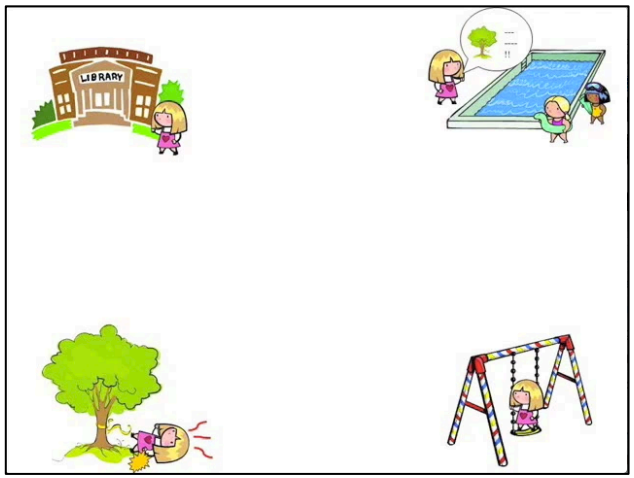


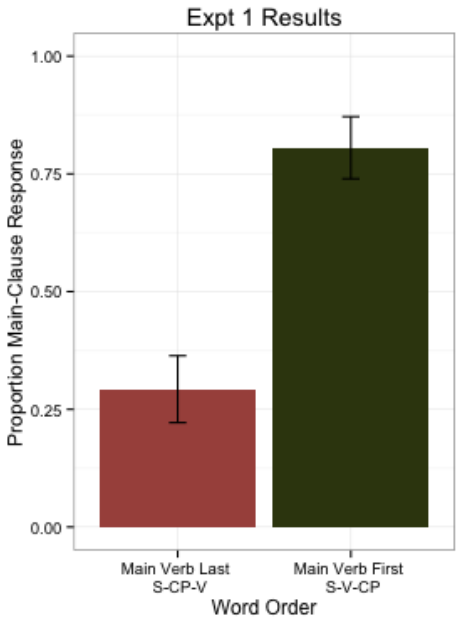


Bengali

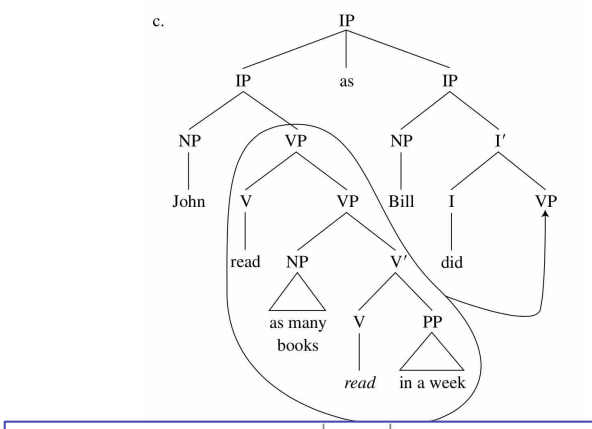
Verb-medial
Emily kothae bollo [*je nijeke bætha korlo*]
Emily where said that self hurt did

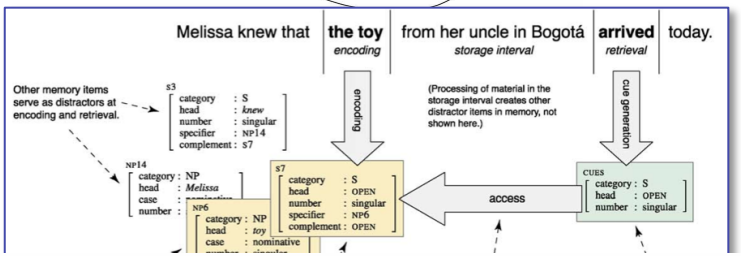
Verb-final
Emily kothae [*nijeke bætha korlo bole*] *bollo*
Emily where self hurt did that said
'Where did Emily say that she hurt herself?'

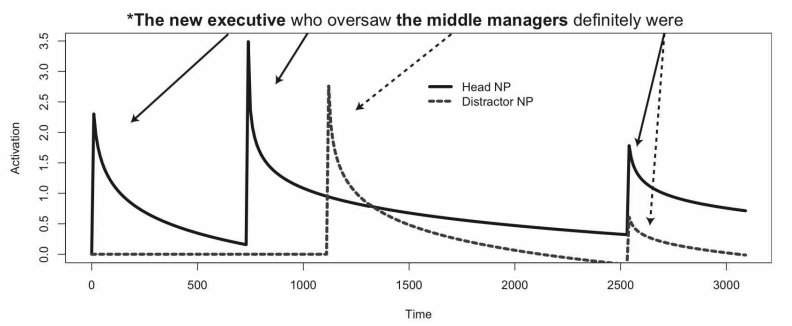


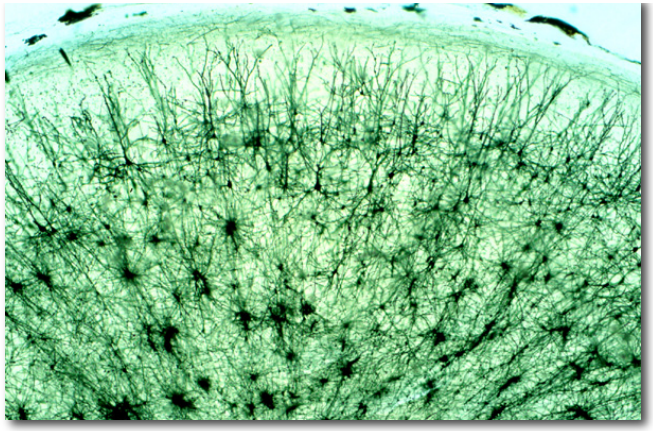


Dustin Chacón, Mashrur Imtiaz









structural constraints

(left-right) order

time

resource limitations

memory architecture and access

non-discrete units

etc.

Locality

Structural constraints exist, and are deployed rapidly.

But they are not reducible to constraints of time or memory resources

3. Slow Prediction

Linking linguistic knowledge and real-world knowledge

Ellen Lau
Maryland

Wing Yee Chow
UC London

Shota Momma
Maryland

What do bees do?

Generating predictions

The gardener talked as the barber trimmed the ___mustache___ *…*

The barber talked as the gardener trimmed the ___hedge___ *…*

I'm not going to solely blame all of man's activities on changes in climate

9/30/08

I'm not one to attribute every activity of man to climate change

Role-reversal errors in comprehension are real, but how pervasive are they?

(i) syntax-free interpretive mechanism
(ii) recovery from error
(iii) when intended meaning known

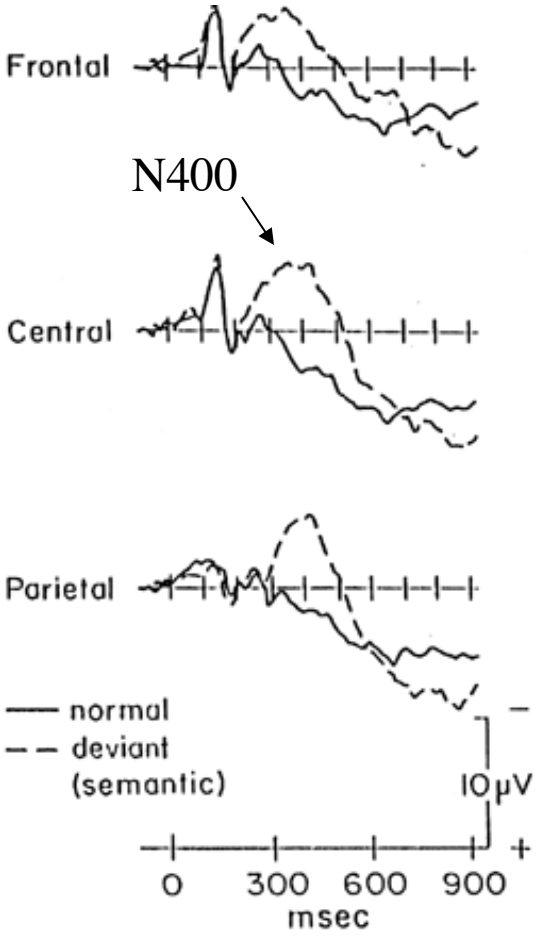
10/02/08

Electrophysiology of Sentence Comprehension

N400

- Semantic anomaly

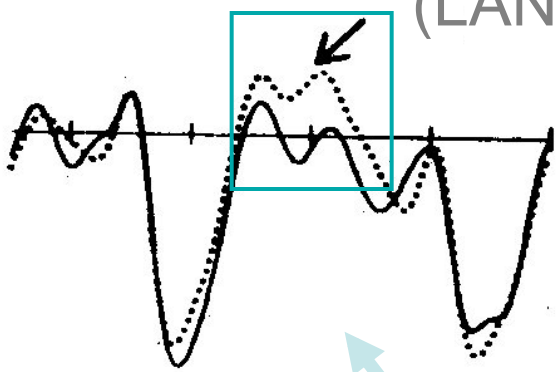
I drink my coffee with cream and sugar
I drink my coffee with cream and socks

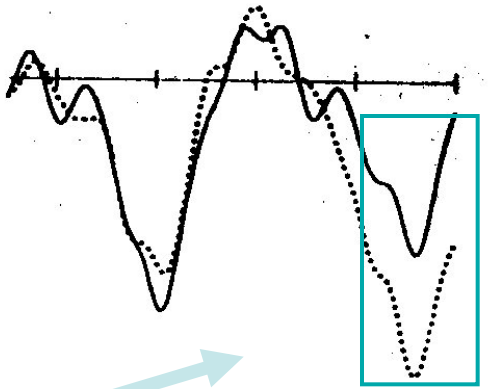


Kutas & Hillyard (1980)

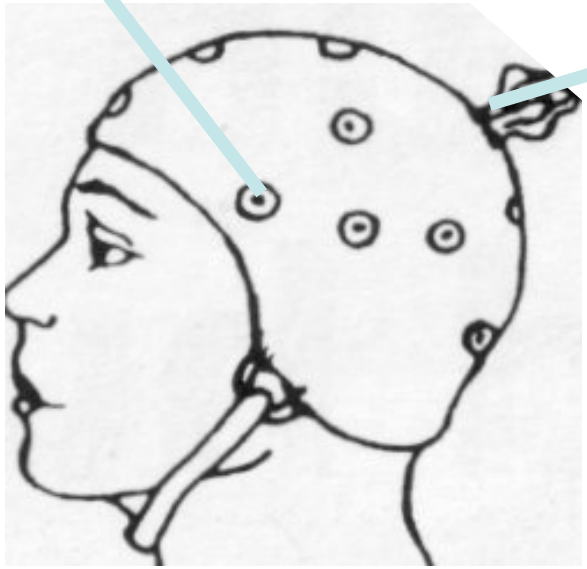
Electrophysiology of Sentence Comprehension

Left Anterior Negativity (LAN)



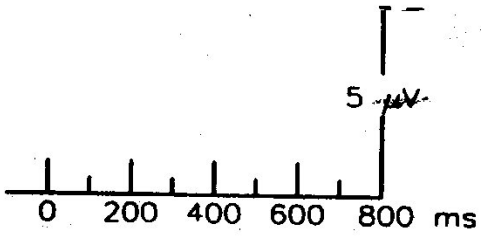


P600



he mows

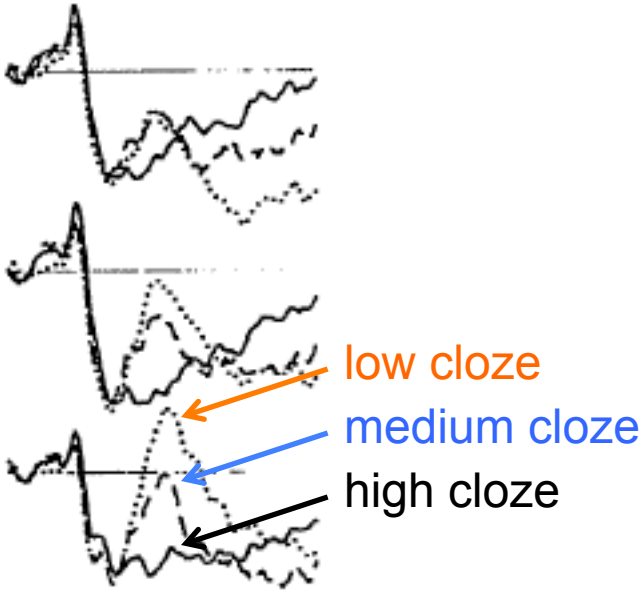
he *mow



Where do we usually see an N400 effect?

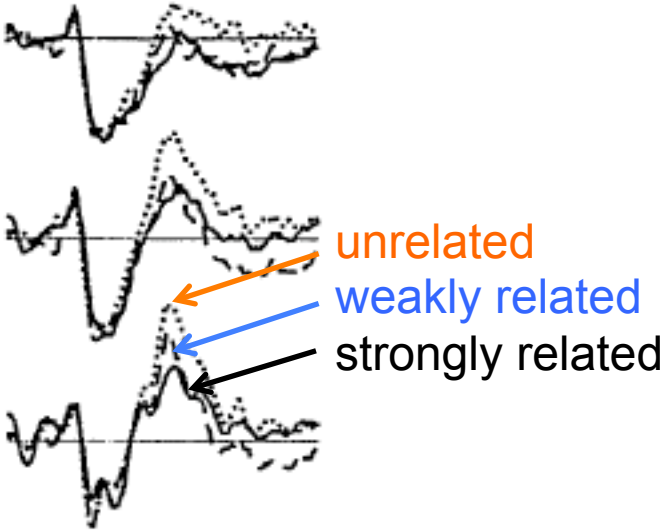
Cloze Probability

- He was stung by a …
  … **bee** / **wasp** / **fly**

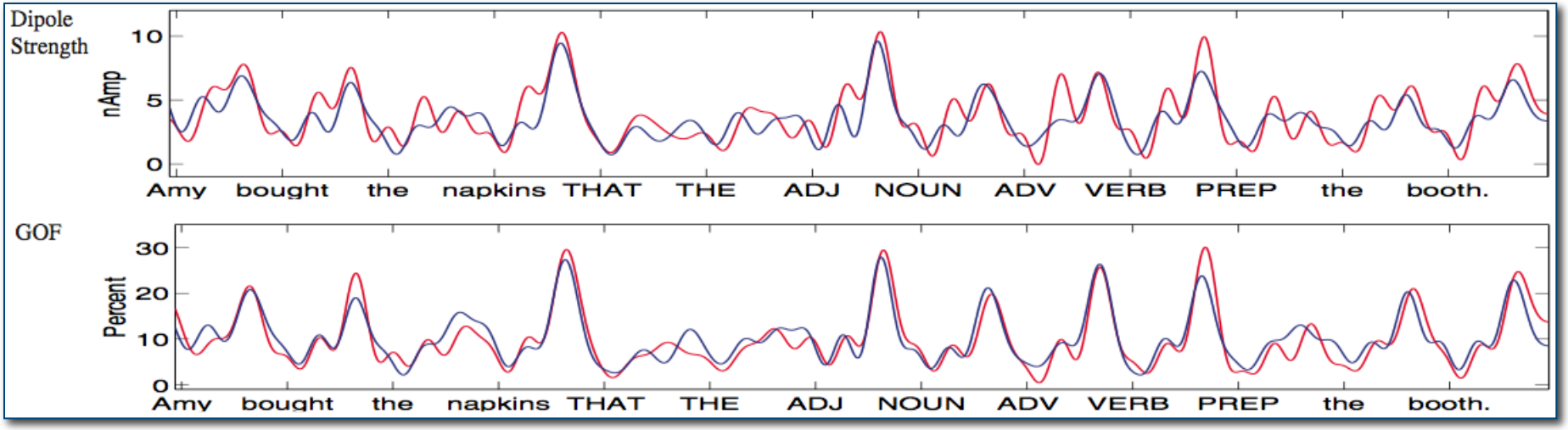


Priming

- doctor – **NURSE**
- table – **NURSE**

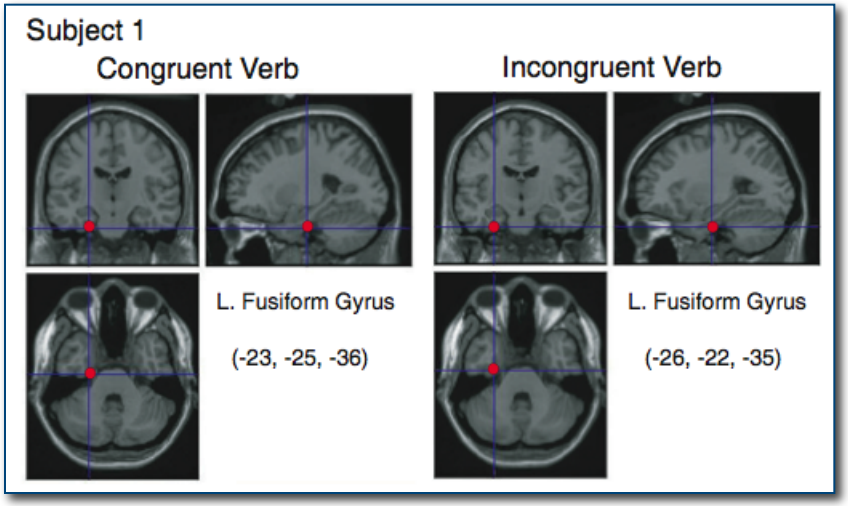


Kutas (1993)



Amy bought the napkins that the café manager diligently folded in the booth.
Amy bought the napkins that the café manager diligently baked in the booth.

MEG counterpart of N400 at successive word positions in sentence comprehension.



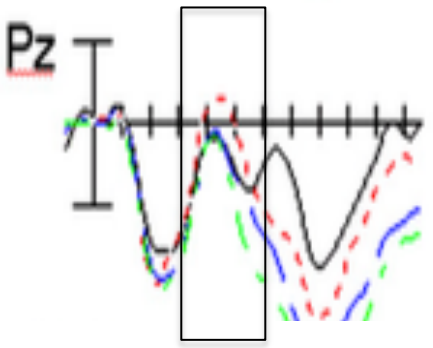
(Yeung *et al.*, 2004)

Failure to predict?

Some aspects of context fails to influence N400

Argument role-reversal

At the breakfast the boy would eat …
?At the breakfast the egg would eat …



No N400 diff. b/w Black & Blue.

Kuperberg et al. 2007; Kim & Osterhout 2005; Kolk et al., 2003; Chow & Phillips 2013; van Herten et al., 2006; Bornkessel & Schlesewsky, 2006; Kolk & Chwilla, 2007; Hoeks et al., 2004)

Minimal Reversal

Japanese

- Role reversal by case-marker reversal (NOM-ACC)
- Two-word sentences (one argument dropped).

Bee-**NOM** STING

Fish-**ACC** CATCH

Scholar-**NOM** STUDY

God-**ACC** WORSHIP

Minimal Reversal

Japanese

- Role reversal by case-marker reversal (NOM-ACC)
- Two-word sentences (one argument dropped).

?Bee-**ACC** STING

?Fish-**NOM** CATCH

?Scholar-**ACC** STUDY

?God-**NOM** WORSHIP

Minimal Reversal

Japanese

- Role reversal by case-marker reversal (NOM-ACC)
- Two-word sentences (one argument dropped).

Canonical vs. Role-reversed

Bee-**NOM** STING vs. ? Bee-**ACC** STING

Fish-**ACC** CATCH vs. ? Fish-**NOM** CATCH

Scholar-**NOM** STUDY vs. ? Scholar-**ACC** STUDY

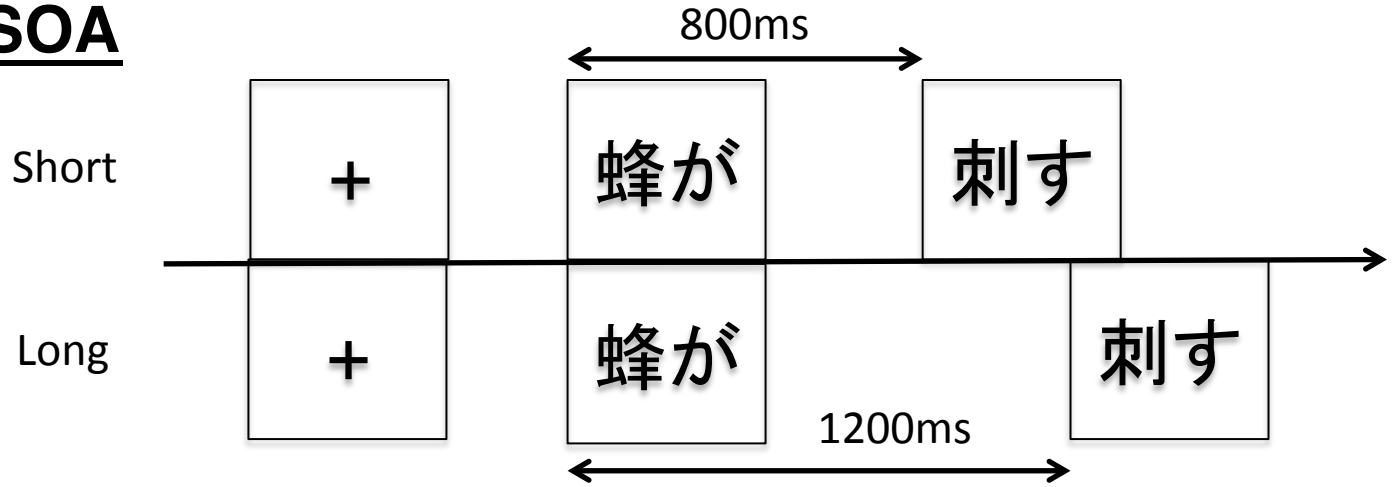
God-**ACC** WORSHIP vs. ? God-**NOM** WORSHIP

Design

Factor 1: Plausibility

Plausible	Implausible
Bee-NOM sting	Bee-ACC sting
Fish-ACC catch	Fish-NOM catch

Factor 2: SOA



Method details

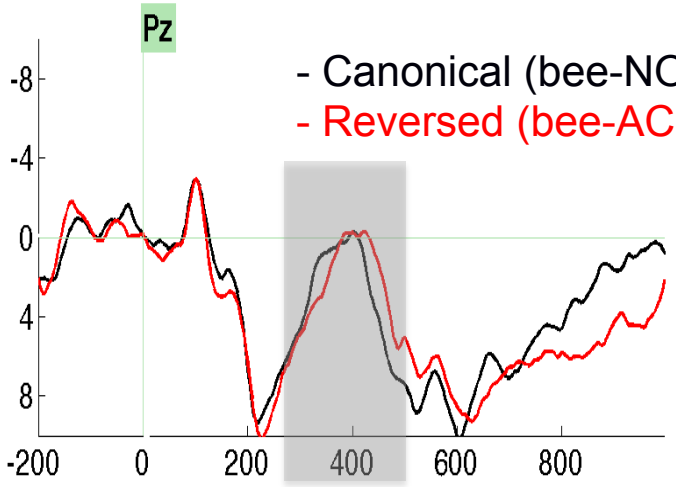
N =24, all right handed native Japanese speakers

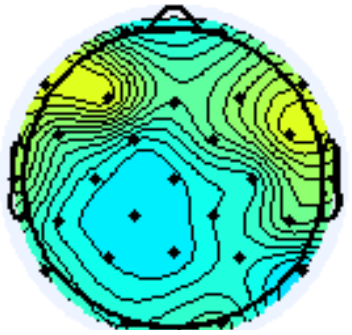
Delayed plausibility judgment;

160 Experimental (40 per condition) + 160 'filler' sentences; Latin-Square

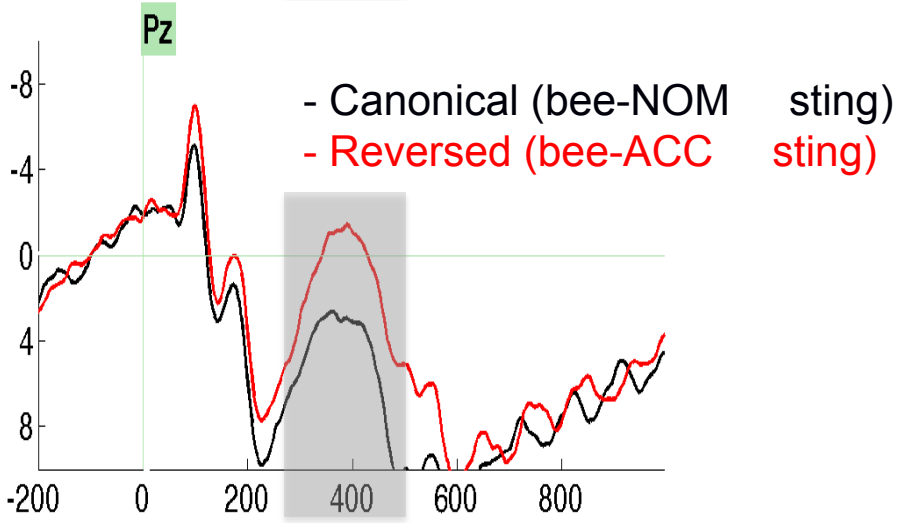
Result – Verb

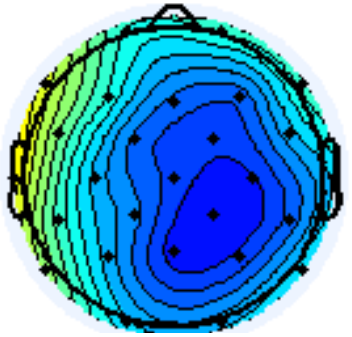
Short SOA



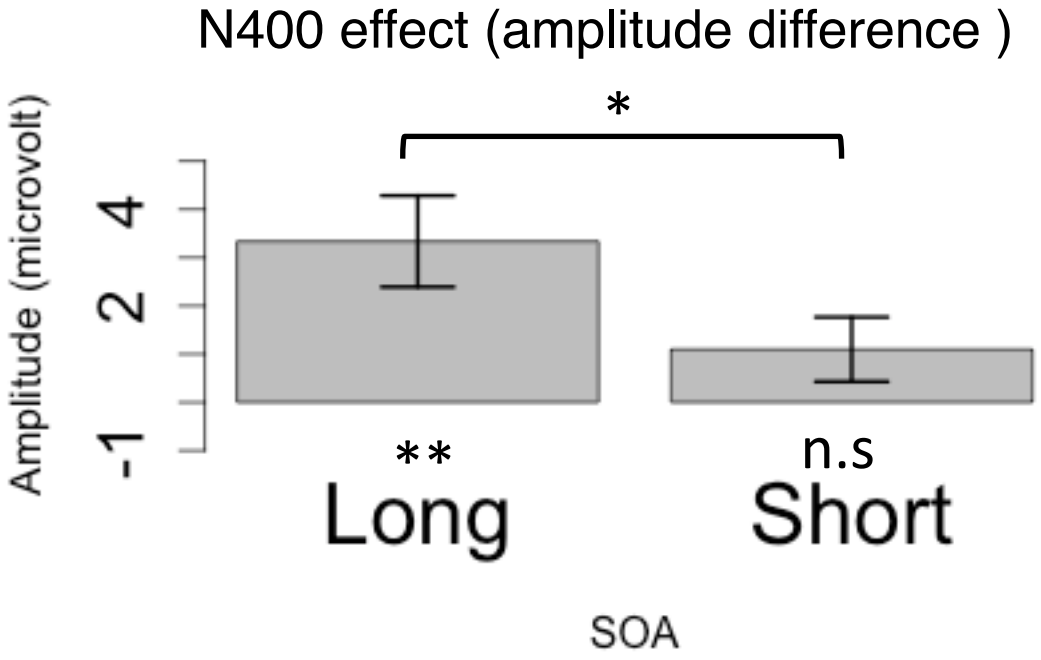


Long SOA





Result – Verb



Average midline electrodes

Result – Control

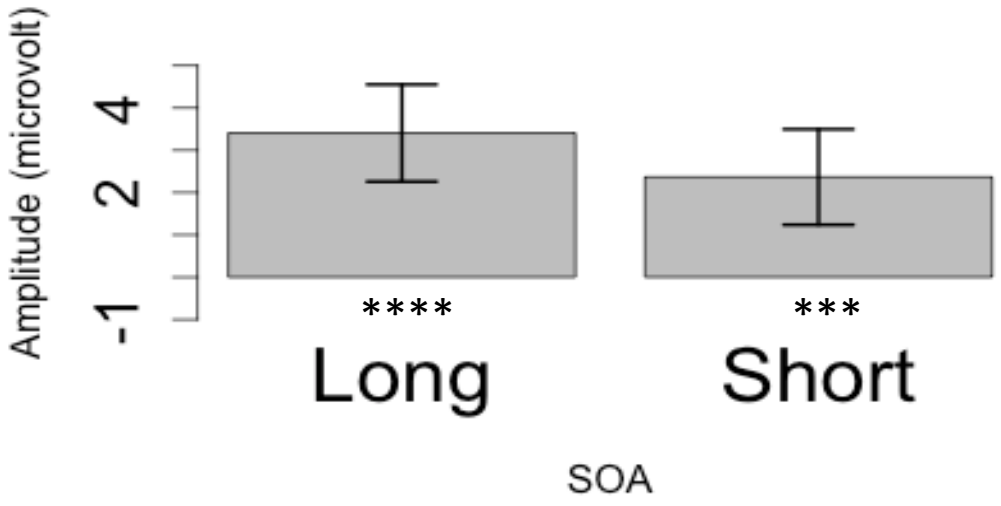
Plausible

Turtle-NOM swim
Apple-ACC eat

Implausible

Turtle-NOM gets-cold
Apple-ACC shave

N400 effect (amplitude difference)



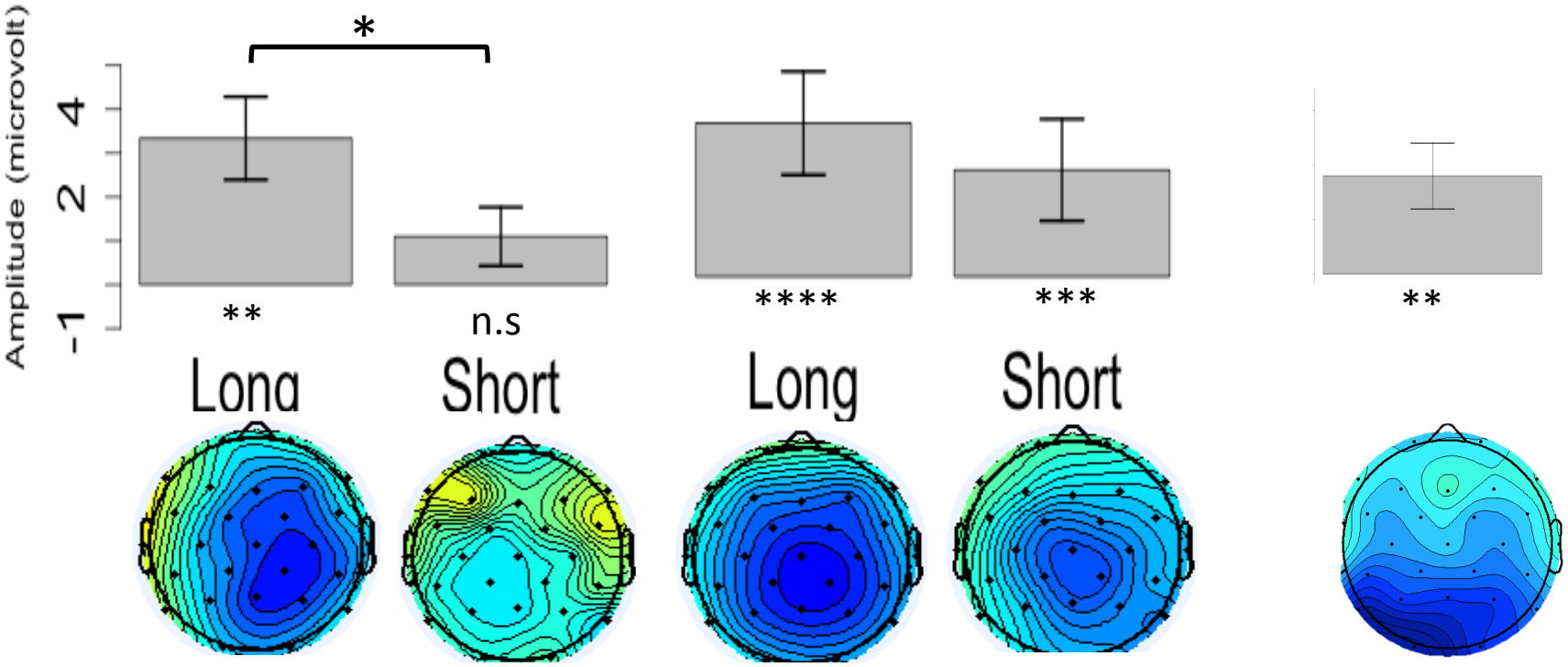
Verb

Noun

Bee-NOM sting vs. Bee-ACC sting

Apple-ACC eat vs. Apple-ACC shave

NOM vs. ACC



Whatever the source of this difference, it is due to something that happens *before* the verb.

-> Slow prediction, not failure to predict/ N400's failure to reflect prediction

This is probably *not* due to people's inattentiveness to the case morph.

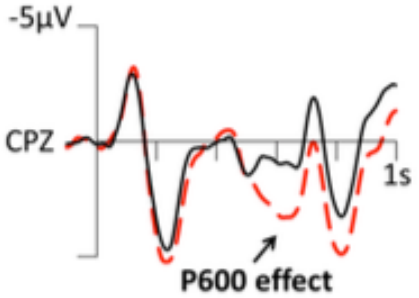
Same pattern in verb-final structure in Chinese:

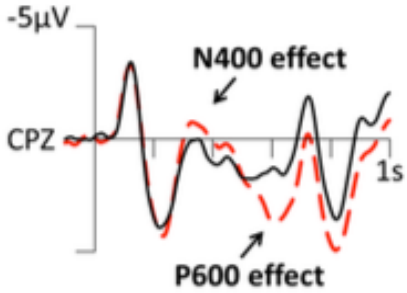
Last week *policeman BA suspect arrest…*

Last week *suspect BA policeman arrest…*

Policeman BA suspect ***ZAI*** ***last week*** *arrest…*

Suspect BA Policeman ***ZAI*** ***last week*** *arrest…*

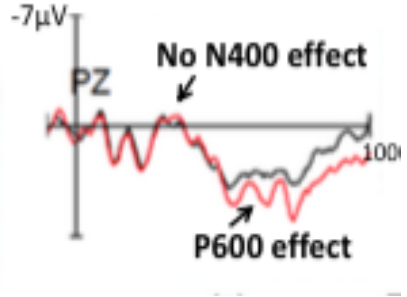




Same cloze contrast, different source:

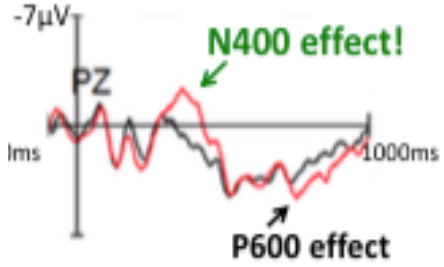
… which customer the waitress had served… (25.4%)

… which waitress the customer had served… (zero)



… which illustrator the author had hired… (27.7%)

… which readers the author had hired… (zero)



Chow et al. (2015 ab); see Kukona et al. 2011 for eye-tracking evidence.

Prediction as a memory search

Prediction = memory search

Lexical prediction ~= semantic memory search problem

Given *bee-ACC…*

what does it takes to predict *swat* (or whatever appropriate verb) instead of *sting*?

- Search for ***a predicate that typically takes bee as a patient***.
- Can we spot this item in semantic memory in one direct step? – probably not.

Encoding and accessing structured information in memory

"Can you find the man in the stripy sweater?"

Spreading activation as search space reduction

Spotting the right items

Sting

Pain

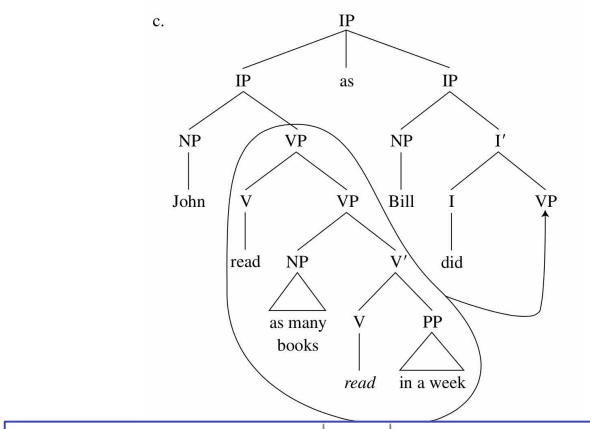
Buzz

Honey

Swat

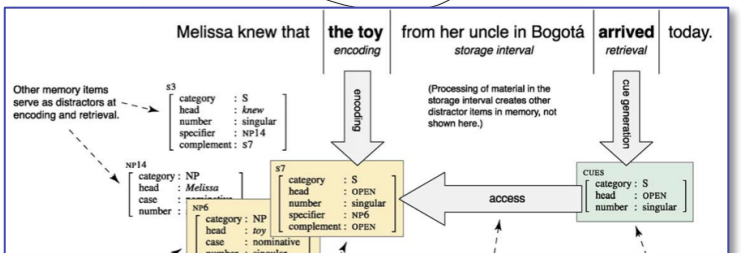
Insect

Hive

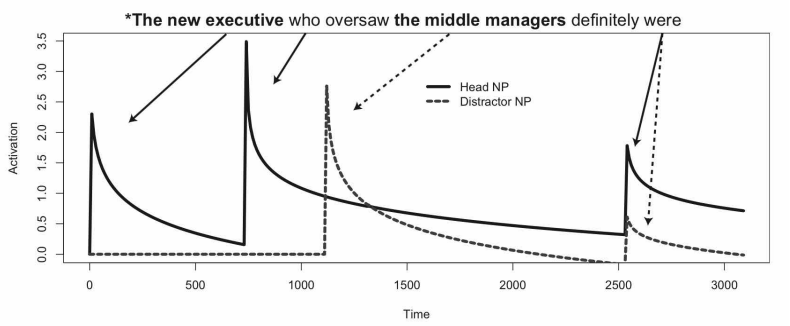


structural constraints

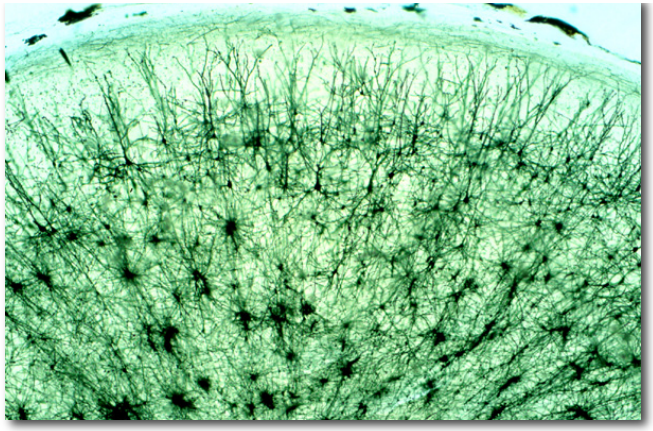
(left-right) order



time



resource limitations



memory architecture and access

non-discrete units

etc.

Summary

One-system architecture

No distinct “processor”

Multiple levels of analysis

Memory

1. Tension between levels of analysis
2. Reductionism
3. Prediction & memory

papers at ling.umd.edu/colin

Class 2

Relating Parsing and Grammar:

20 years on

Colin Phillips

Right Association in Parsing and Grammar*

Colin Phillips, *Massachusetts Institute of Technology*

1. Introduction

It is straightforward enough to show that sentence parsing and grammaticality judgements are different. There are sentences which are easy to parse but ungrammatical (e.g. *that-trace* effects), and there are sentences which are extremely difficult to parse, but which may be judged grammatical given appropriate time for reflection (e.g. multiply center embedded sentences). This classic argument shows that parsing and grammar are not identical, but it tells us very little about just how much they have in common. The goal of this paper is to point to one principle which may account for a number of effects of ambiguity in both domains.

1995

ABSTRACT

The aim of this thesis is to argue for the following two main points. First, that grammars of natural language construct sentences in a strictly left-to-right fashion, i.e. starting at the beginning of the sentence and ending at the end. Second, that there is no distinction between the grammar and the parser.

In the area of phrase structure, I show that the left-to-right derivations forced by the principle *Merge Right* can account for the apparent contradictions that different tests of constituency show, and that they also provide an explanation for *why* the different tests yield the results that they do. Phenomena discussed include coordination, movement, ellipsis, binding, right node raising and scope.

I present a preliminary account of the interface of phonology and morphology with syntax based on left-to-right derivations. I show that this approach to morphosyntax allows for a uniform account of locality in head movement and clitic placement, explains certain directional asymmetries in phonology-syntax mismatches and head movement, and allows for a tighter connection between syntactic and phonological phrases than commonly assumed.

In parsing I argue that a wide range of structural biases in ambiguity resolution can be accounted for by the single principle *Branch Right*, which favors building right-branching structures wherever possible. Evidence from novel and existing experimental work is presented which shows that Branch Right has broader empirical coverage than other proposed structural parsing principles. Moreover, Branch Right is not a parsing-specific principle: it is independently motivated as an economy principle of syntax in the chapters on syntax.

The combination of these results from syntax and parsing makes it possible to claim that the parser and the grammar are identical. The possibility that the parser and the grammar are identical or extremely similar was explored in the early 1960s, but is widely considered to have been discredited by the end of that decade. I show that arguments against this model which were once valid no longer apply given left-to-right syntax and the view of the parser proposed here.

1996

Derivational Order in Syntax: Evidence and Architectural Consequences*

COLIN PHILLIPS & SHEVAUN LEWIS
University of Maryland
{colin, shevaun}@umd.edu

Most formal syntactic theories propose either that syntactic representations are the product of a derivation that assembles words and phrases in a bottom-to-top and typically right-to-left fashion, or that they are not constructed in an ordered fashion. Both of these views contrast with the (roughly) left-to-right order of structure assembly in language use, and with some recent claims that syntactic derivations and real-time structure-building are essentially the same. In this article we discuss the mentalistic commitments of standard syntactic theories, distinguishing *literalist*, *formalist*, and *extensionalist* views of syntactic derivations. We argue that existing evidence favors the view that human grammatical representations are the product of an *implementation dependent* system, i.e., syntactic representations are assembled in a consistent order, as claimed by grammatical models that are closely aligned with real-time processes. We discuss the evidence for left-to-right syntactic derivations, and respond to critiques of a proposal that the conflicts between the results of constituency diagnostics can be explained in terms of timing.

2013

J Psycholinguist Res (2015) 44:27–46
DOI 10.1007/s10936-014-9329-z

Aligning Grammatical Theories and Language Processing Models

Shevaun Lewis · Colin Phillips

Published online: 19 November 2014

Abstract We address two important questions about the relationship between theoretical linguistics and psycholinguistics. First, do grammatical theories and language processing models describe separate cognitive systems, or are they accounts of different aspects of the same system? We argue that most evidence is consistent with the one-system view. Second, how should we relate grammatical theories and language processing models to each other?

Keywords Parsing · Grammatical theories · Abstraction · Cognitive architecture of language

2015

2(+)-system accounts

- Relation between linguistic **knowledge** and **real-time mechanisms**

  Standard view: there are distinct cognitive language systems

- "… ongoing speech behavior does not utilize a linguistic grammar. [*evidence from various studies*] … invalidates any model for speech recognition that attempts to directly incorporate grammatical rules …" (Bever 1970; also Fodor et al. 1974)

- Motivations:
  (i) Arguments about reversing grammatical derivations
  (ii) Slaying the DTC dragon [**D**erivational **T**heory of **C**omplexity]
  (iii) Comprehension ≠ Production
  (iv) Judgments take time, but comprehension is fast
  (v) Grammars lack some key features of parsing systems
  (vi) People do dumb stuff

1-system account

- Knowledge of language is knowledge of how to construct representations in real time.

- The mental grammar is **implementation dependent** – there is only one way to build a given representation. [Phillips & Lewis 2013]

  Main Claim: there is no evidence of separate grammatical computations that occur on a different time scale than normal parsing and production.

  (got evidence? let me know!)

- Revising past claims (Phillips 1996, 2003, 2004)

  (i) "The parser is the grammar" ✖
  (ii) Grammatical derivations are strictly left-right ✖
  (iii) Real-time processes are grammatically 'infallible' ✖
  (iv) Concern with economy metrics ('branch right') ✖

Grammatical Analysis

- I have done little new work in this area since ~2000.
- Reason 1: practical
- Reason 2: theoretical
- Most findings in grammatical theory require no commitment to order
- Reductionism

- Related arguments

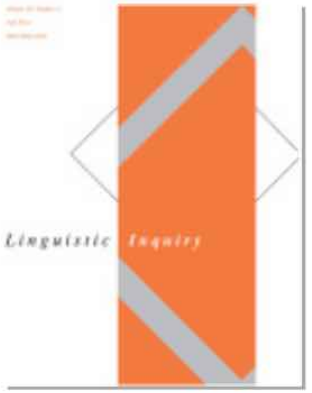
  - Schlenker (1999): German adjective inflection
  - Boeckx (1999): islands
  - Richards (2000): 'subjacency tax' effects in Japanese
  - Guimaraes (2003): syntactic amalgams
  - McGinnis & Gerdts (2003): Kinyarwanda applicatives
  - Barker & Shan (2006): crossover phenomena
  - Chung et al. (2006): licensing 'sprouting'
  - Choi & Yoon (2006): Japanese/Korean argument clusters
  - Kempson, Cann et al. (2001/2005): varied phenomena [*Dynamic Syntax*]
  - Steedman (2000, etc.): various phenomena [*Comb. Categorial Grammar*]

Subject Islands, Reconstruction, and the Flow of the Computation

Valentina Bianchi, Cristiano Chesi

Linguistic Inquiry, Volume 45, Number 4, Fall 2014, pp. 525-569 (Article)

Published by The MIT Press



Grammar in Comprehension

- Fast/slow mis-alignments
- Implementing grammatical constraints in memory architecture
- Harder case #1: head-finality
- Harder case #2: incremental semantic interpretation

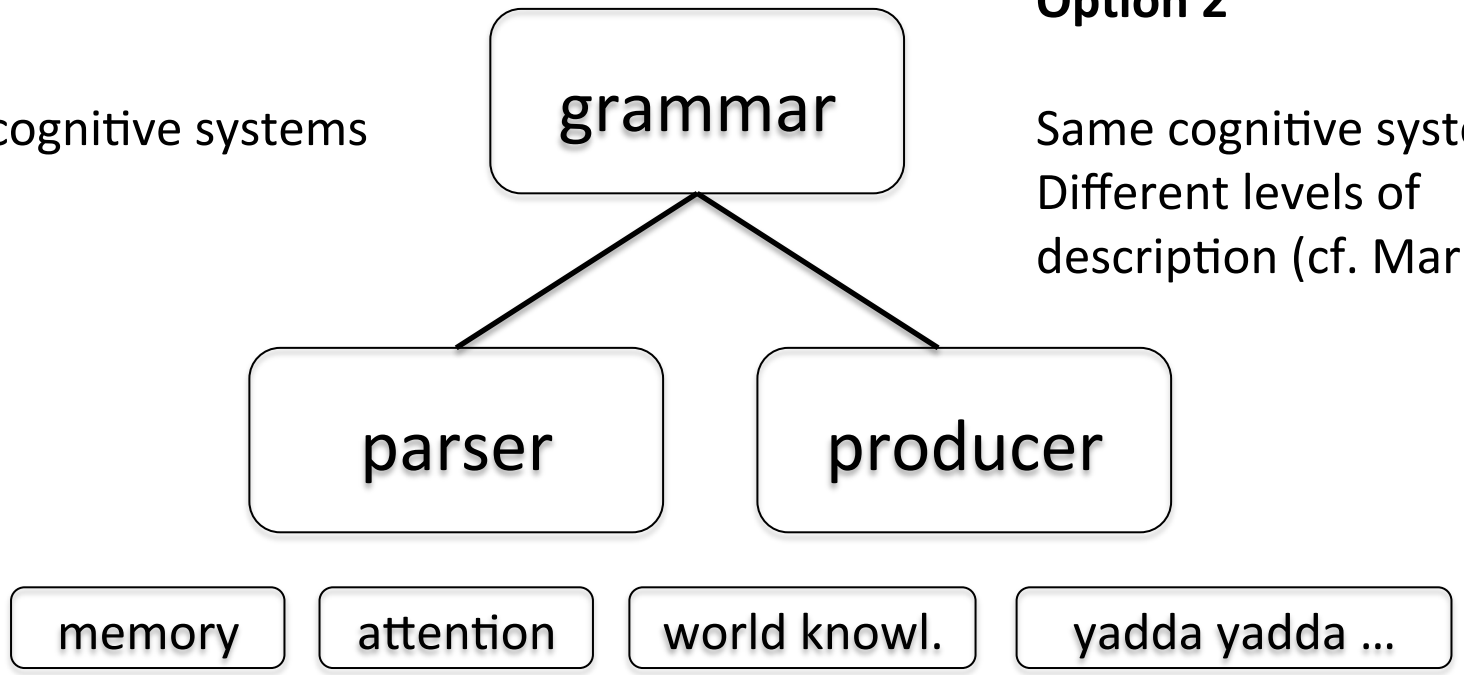
Comprehension-Production

- Does structure building follow same steps in comprehension and production?
- Does predictive structure building in comprehension mimic the language production system?
- Do comprehension & production imply distinct cognitive systems?

But what are we aligning?

Option 1

Distinct cognitive systems



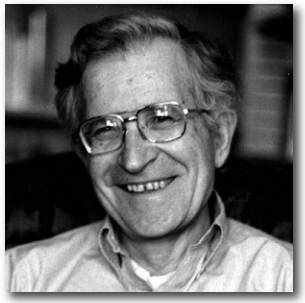
Option 2

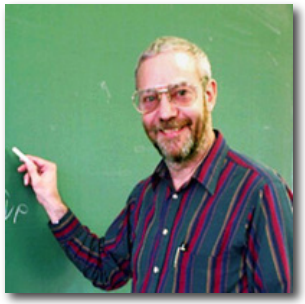
Same cognitive system;
Different levels of description (cf. Marr 1982)

Psycholinguist 1: “I’ve always assumed this was a discussion about the separate systems view.”

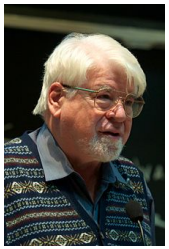
Psycholinguist 2: “We’ve mostly assumed that levels-of-description view.”

Close Alignment is not Desirable





"It has sometimes been argued that linguistic theory must meet the empirical condition that it account for the ease and rapidity of parsing. But parsing does not, in fact, have these properties. […] In general, it is not the case that language is readily usable or 'designed for use.'" (Chomsky & Lasnik, 1993, p. 18)



"we understand everything twice"
(Townsend & Bever, 2001)

"…the language comprehension system creates representations that are 'good enough' (GE) given the task that the comprehender needs to perform." (Ferreira & Patson, 2007, p. 71)

"Two interpretive systems for natural language." (Frazier)

Close Alignment is Desirable

A grammatical model that is easily embedded in a real-time parsing/production system is clearly preferable.

Some Desiderata:

1. Incremental elaboration of partial structures/interpretations

2. Computationally explicit

3. Alignment of constraints on well-formedness and constraints on use


Steedman


Sag


Chesi


Bresnan


Kempson


Kempen


Hawkins

Mentalistic Commitments of Standard Theories

These are clear mentalistic hypotheses, empirically testable, based on what people do

A. Literalist

Derivation is a literal sequence of mental operations that language users can, in principle, carry out. (e.g., Townsend & Bever 2001)

B. Formalist

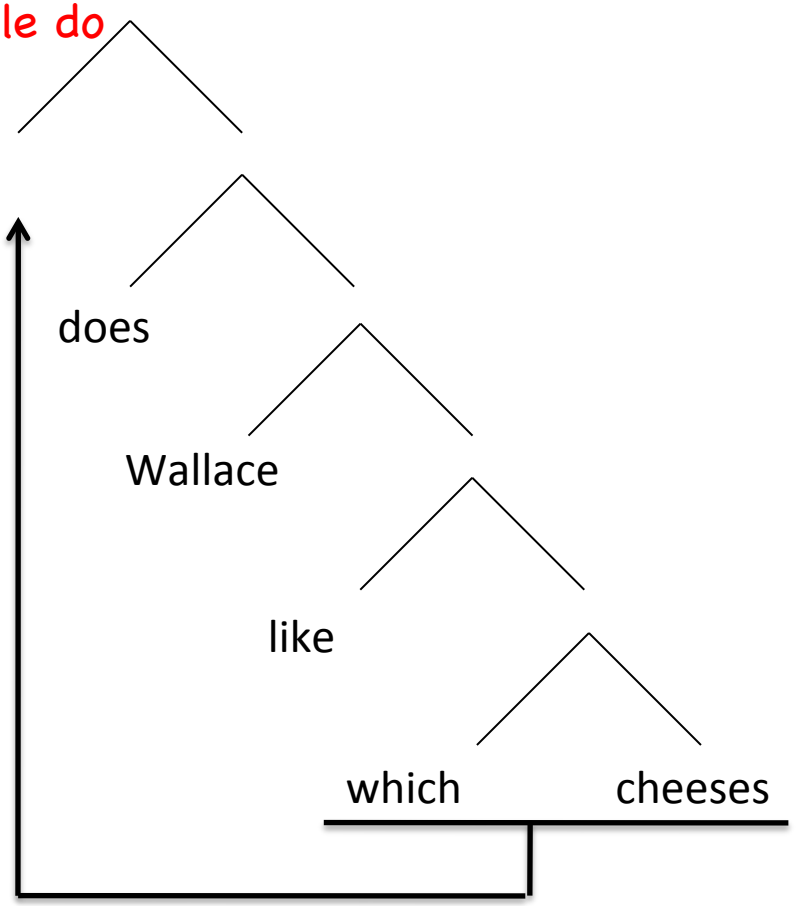
Speakers do construct the set of representations in a derivation (at least, LF, PF), but these respect a *formal* ordering relation; rather than a strict *temporal* order

C. Extensionalist

Derivations are not hypotheses about sequences of mental representations, on any time scale. Accountable only to successful partitioning of ±grammatical sentences.

(i) 'Pragmatic' extensionalist

(ii) 'Principled' extensionalist



Beyond the reach of most empirical evidence. Components of theory have no independent status as cognitive objects. ? 'Efficiency', 'economy', 'learning' ?

Motivations for Standard View

(i) Arguments about reversing grammatical derivations [In Aspects model]

(ii) Slaying the DTC dragon [**D**erivational **T**heory of **C**omplexity]

(iii) Judgments can be slow

(iv) Grammars lack key features of parsing systems

(v) Grammatical evidence for bottom-up derivations

(a) thematic structures as 'basic'; (b) cyclicity; (c) endocentricity of phrases < upwards projection; (d) arguments for ordered operations

(vi) Comprehension ≠ Production

More generally: multiple ways of building same representation

(vii) People do dumb stuff

… so the real-time mechanisms don't do what the grammar does

Implementation (In)dependence

- Abstract statements can be very useful!

  e.g., 48 ÷ 6 = 8

  e.g., English wh-phrases are fronted, Chinese has wh-in-situ

- An **Implementation Independent** abstraction can be realized in different ways by multiple lower-level implementations.

- An **Implementation Dependent** abstraction is only ever realized in one way at a less abstract level of description

- Generative grammars traditionally viewed as implementation independent.

  (cf. Neeleman & van de Koot 2009)

  → speakers should be able to build the same representation in multiple ways

Implementation (In)dependence

- An **Implementation Dependent** abstraction is only ever realized in one way at a less abstract level of description

  → our main claim: human grammars are implementation dependent w.r.t. sequencing of structure building operations

- If human grammars are implementation dependent, the extensionalist characterization of grammars is merely a convenient simplification

  ('soft' abstraction; Jackendoff 2002)

- Is there relevant evidence? Less than we'd like.

  a. Speaking & understanding aren't so different (order, representations, …)
  b. Little evidence that same interpretation can be built in different orders
     e.g., reading backwards is difficult-to-impossible
  c. Reanalysis in parsing: more like 're-starting' than about 're-arrangement'

Implementation (In)dependence

- An **Implementation Dependent** abstraction is only ever realized in one way at a less abstract level of description

- Slippery slope - so why not just describe everything in molecular terms?

- No – implementation (in)dependence should be separately assessed at each level of abstraction

- Example – neural encoding of syntactic representations

  On-line construction – too fast to encode via changes in synaptic connectivity
  Long-term memory – plausible only via changes in synaptic connectivity

  → multiple ways of encoding same information neurally

  (… some big caveats to this argument; cf. Jackendoff 2002)

Re-visiting old claims

- "**The Parser is the Grammar**" [PiG] (Phillips 1996)

  No – grammar is one key component of a comprehension system.
  Structure-building system is task neutral: comprehension, production, neither
  Additional mechanisms needed to map to/from external stimuli
  What matters is that there's only one structure-building system

- Grammatical derivations proceed in a **strict left-to-right order** (Phillips 2003)

  No – what matters is that a given sentence is built in only one order
  *Comprehension & production are *roughly* left-to-right*
  Some constructions/languages demand departures, e.g., Japanese head-finality

- Real-time processes should be **grammatically 'infallible'** (Phillips 2004, 2006)

  No – what matters is that there's one way of building the same representation
  Problematic would be evidence of a rough-and-ready 'covering grammar'

What we're not claiming …

- **Comprehension is always successful**

  Grammatically accurate representation ≠ accurate information transfer

  Extreme case: you say something in a noisy bar – I completely misunderstand
  My representation may be perfectly grammatical … just not what you meant

- **Reductionism**

  Claims abound that certain grammatical constraints are epiphenomena – reducible to effects of memory, naturalness, etc.
  (e.g., islands: Erteshik-Shir 1976; Deane 1991; Kluender & Kutas 1993; Hofmeister & Sag 2010 for discussion of reductionist hypotheses, cf. Phillips 2011)

  *Claim that derivations/constraints apply in a fixed order is **not** a reductionist claim*

- Anything about **ambiguity resolution**

Motivations for Standard View

(i) Arguments about reversing grammatical derivations [In Aspects model]

(ii) Slaying the DTC dragon [**D**erivational **T**heory of **C**omplexity]

(iii) Judgments can be slow

(iv) Grammars lack key features of parsing systems

(v) Grammatical evidence for bottom-up derivations

(a) thematic structures as 'basic'; (b) cyclicity; (c) endocentricity of phrases < upwards projection; (d) arguments for ordered operations

(vi) Comprehension ≠ Production

More generally: multiple ways of building same representation

(vii) People do dumb stuff

… mismatch: real-time mechanisms don't do what the grammar does

What is language for?

Organizing and transmitting information between speakers, i.e., *communication*

Language is designed for transmission through serial medium.

Core properties of language should reflect this.

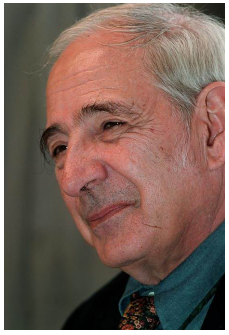
Organizing and transmitting information across time, i.e., *thought*

Language allows us to combine representations flexibly.

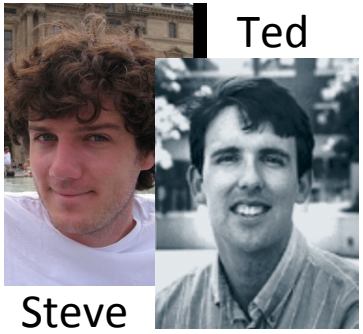
Including propositional embedding.

Language is adaptive even in the absence of communication.

Externalization could be an afterthought.

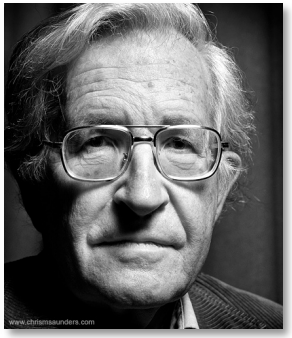

Steedman


Pinker


Steve Ted


Spelke


de Villiers


Chomsky

Standard Grammatical Analysis

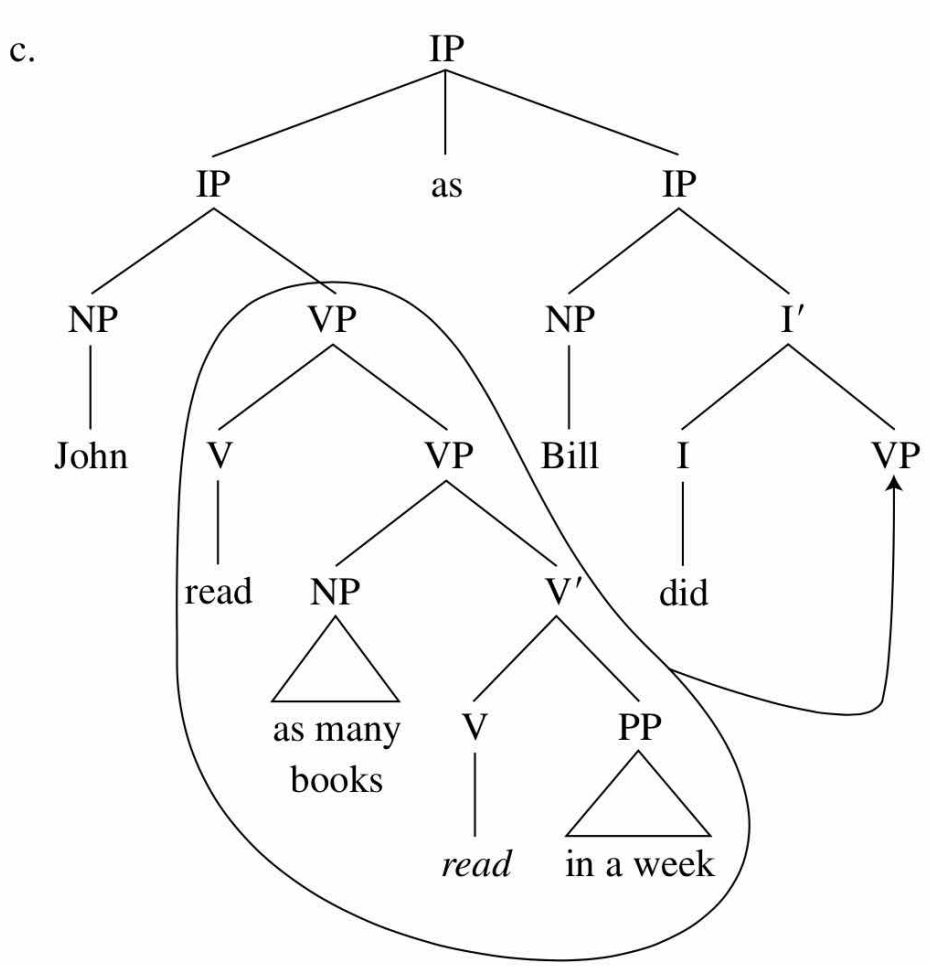
(a.k.a. 'syntactic theory')

Hierarchical groupings of terminals

All elements are discrete symbolic representations

No time dimension

Derivations generally not taken as claims about actual time steps
(Phillips & Lewis 2013)



Default questions:
How acceptable is this sentence?
Why is it so (un)acceptable? Does it violate combinatorial rules? Is it just hard?

… This is a narrow set of questions to ask.

Cognitive Models of Sentence Structure Building

(a.k.a. 'processing theories')

Same questions + many more

Order always matters
Time sometimes matters

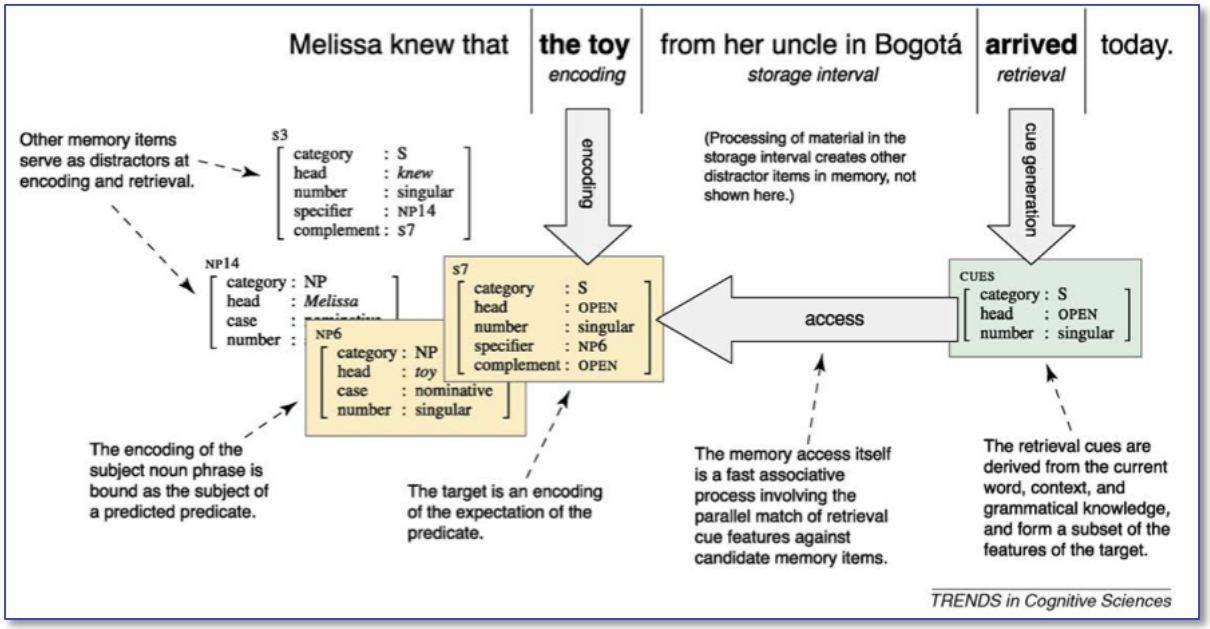
representation =
memory encoding

dependencies =
memory access

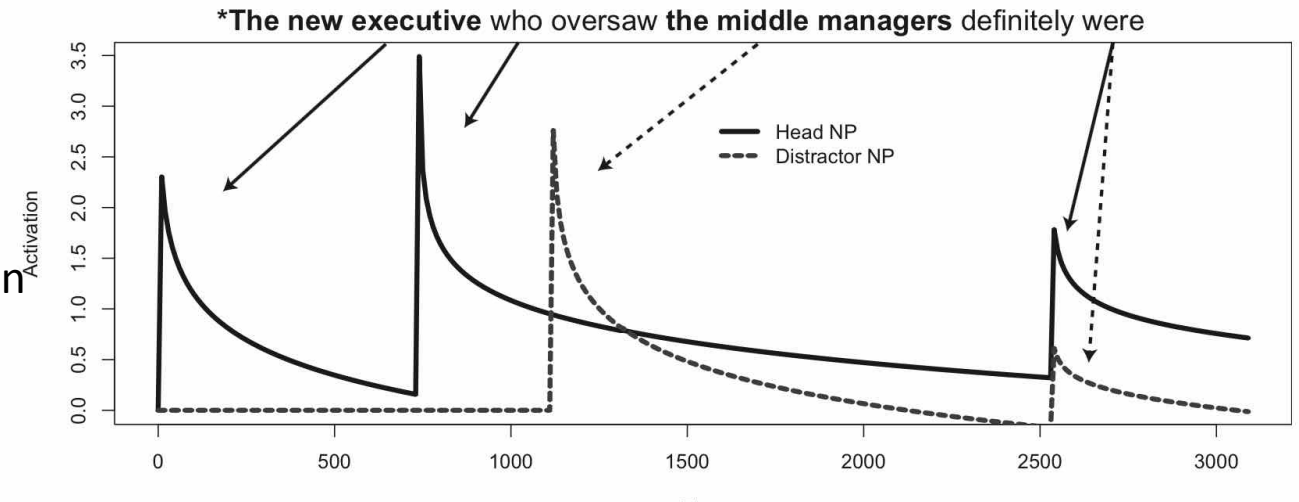
nodes may have varying activation levels

computations may depend on indep. cogn. abilities

comprehension, production work w/ limited information



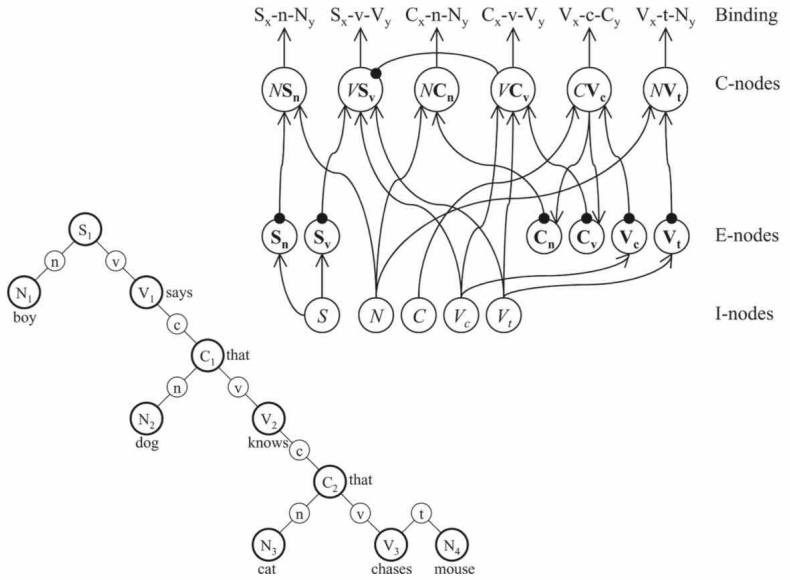
Lewis, Vasishth & Van Dyke 2006

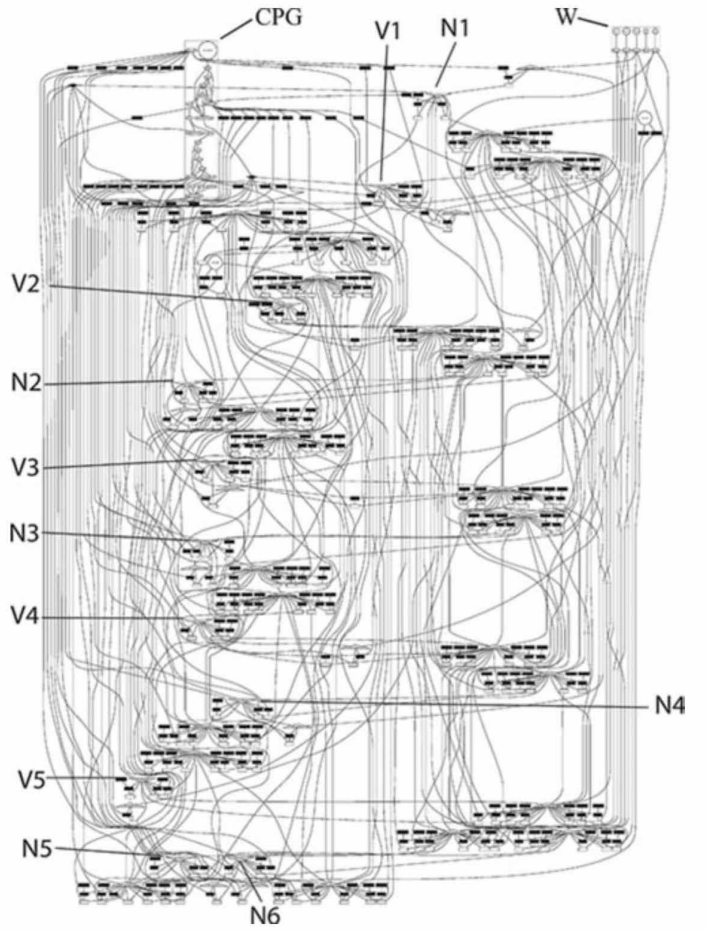


Dillon, Sloggett, Mishler, & Phillips 2013

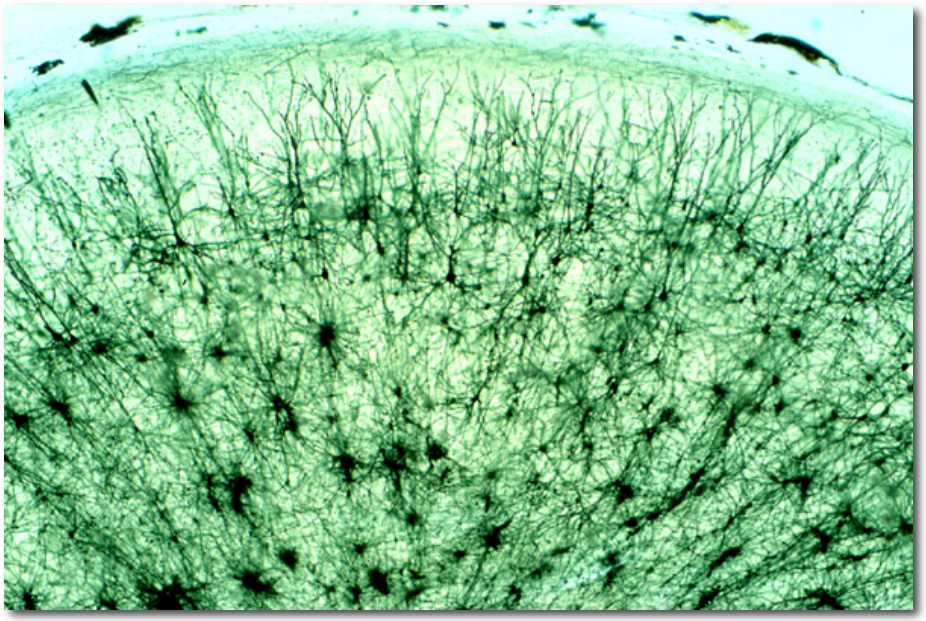
Neural Models of Sentence Structure Building

Different vocabulary / toolkit
Constraints from connectivity & anatomy
Nodes correspond to complex activity pattern

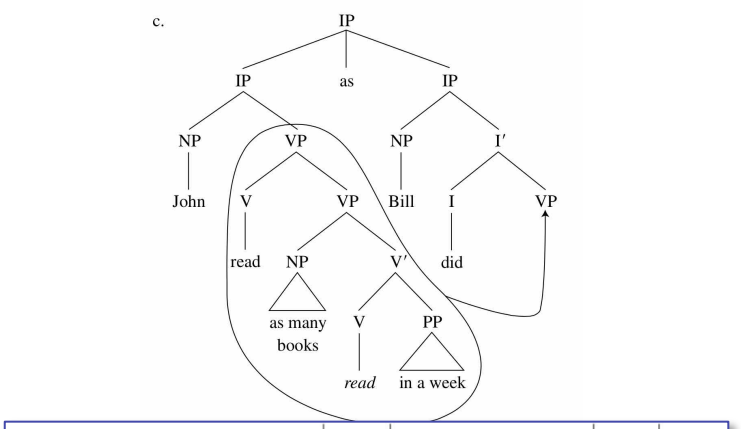




Van der Velde & de Kamps 2006

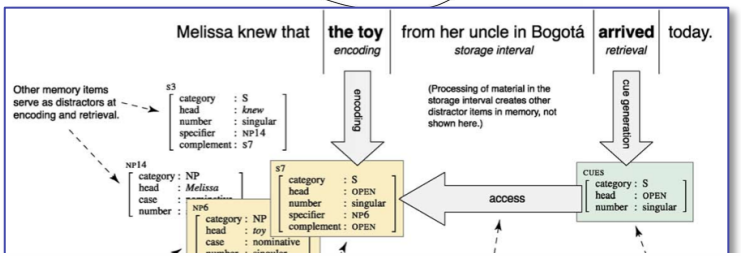


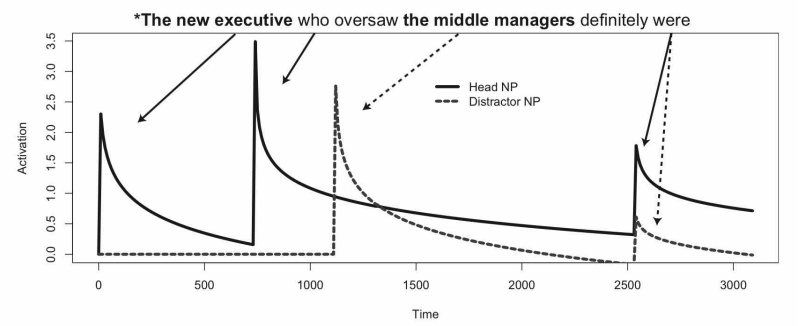
One Language System
Multiple Levels of Analysis



"linguistic"

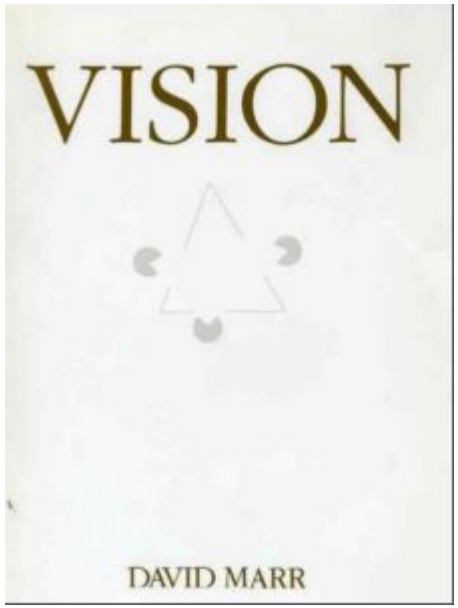
"computational?"



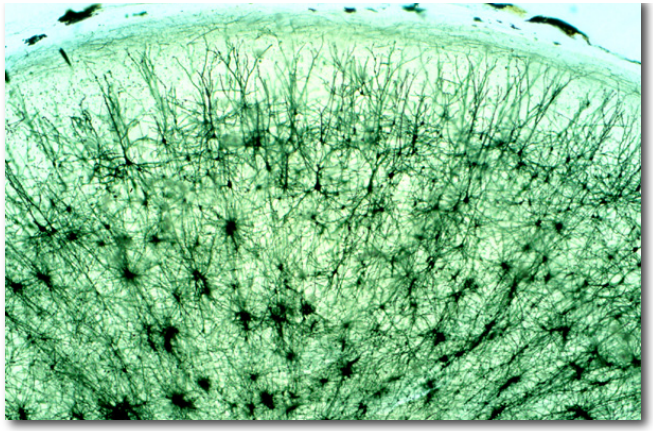


"cognitive"

"algorithmic?"



Marr 1982



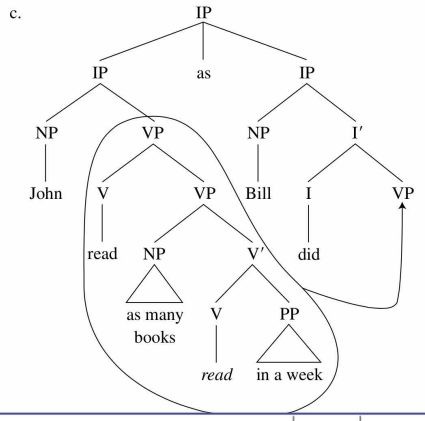
"neural"

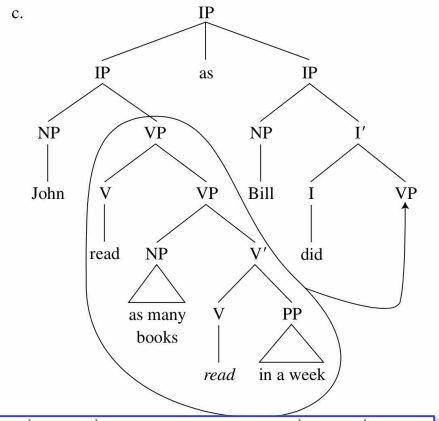
"implementational"

Marr-levels don't really fit language research.

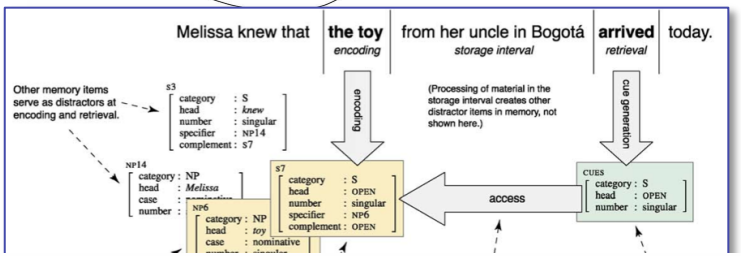
And there aren't really 3 discrete levels.

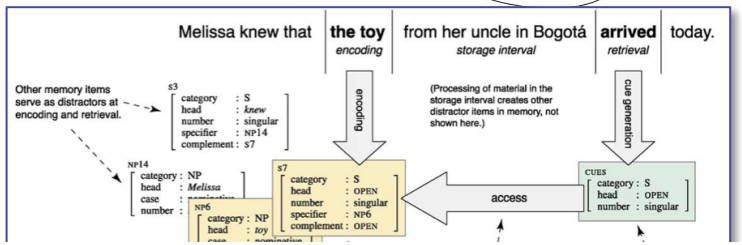
Multiple Language Systems
Multiple Levels of Analysis

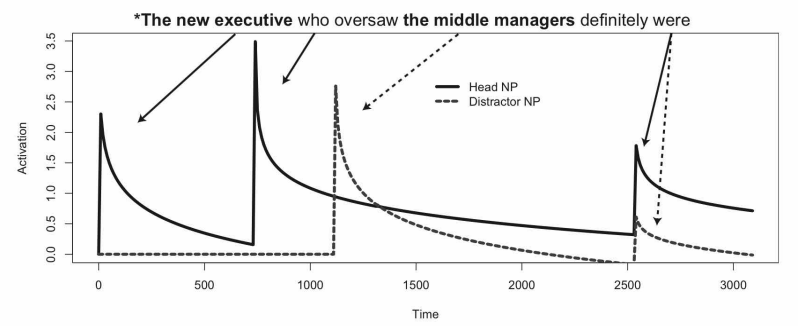


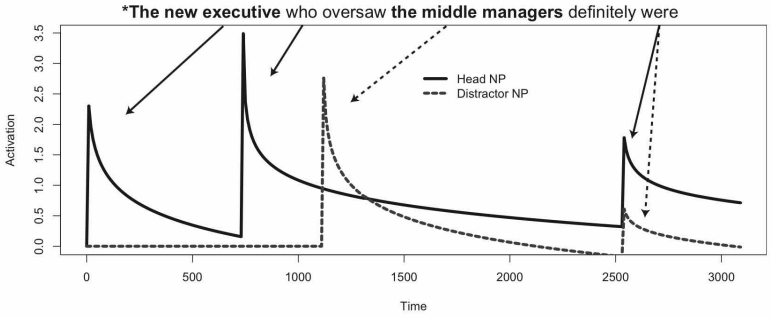


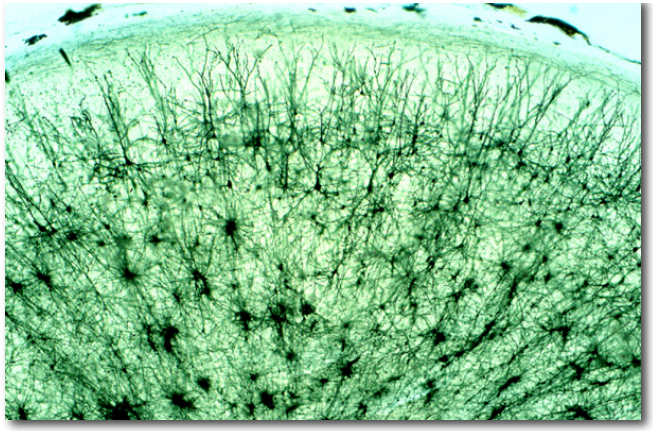
← Grammar
Parser →
Producer

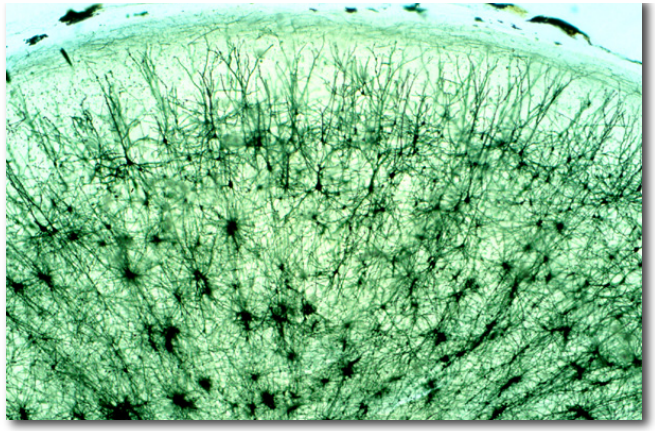


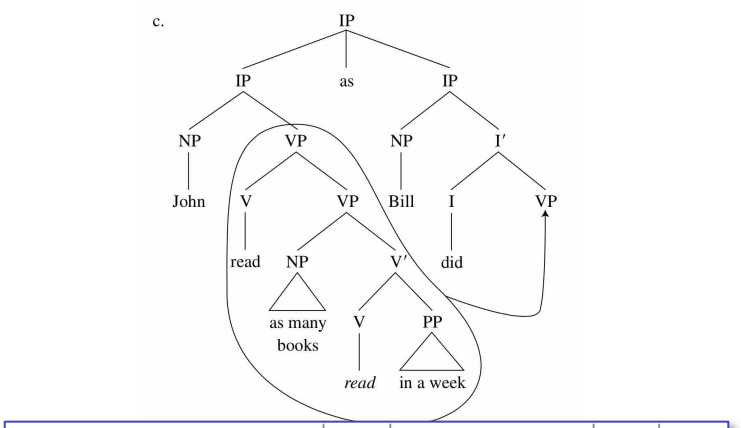






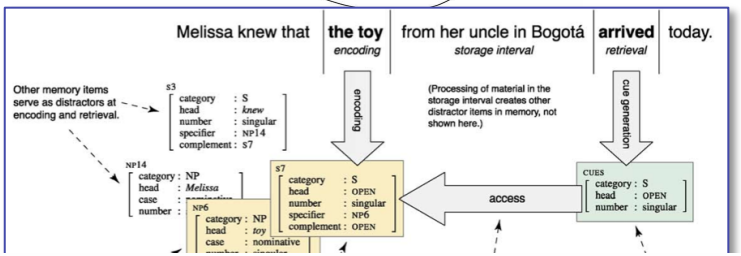






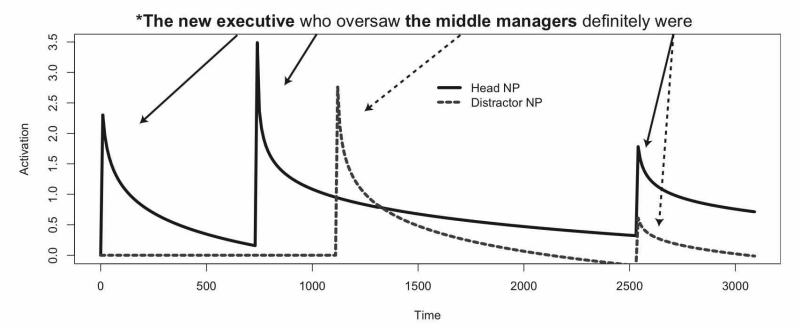
structural constraints

(left-right) order



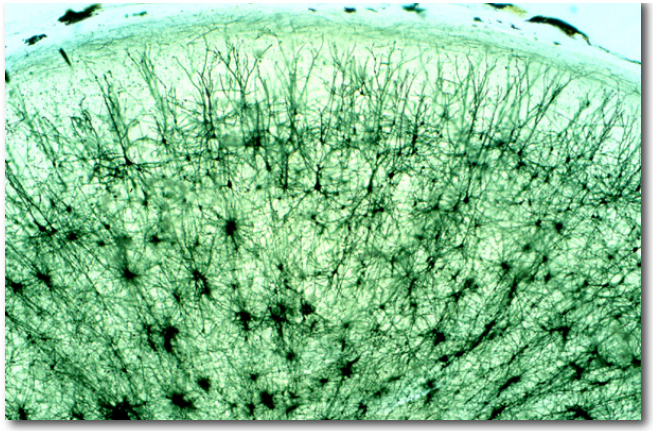
time

Blame Assignment



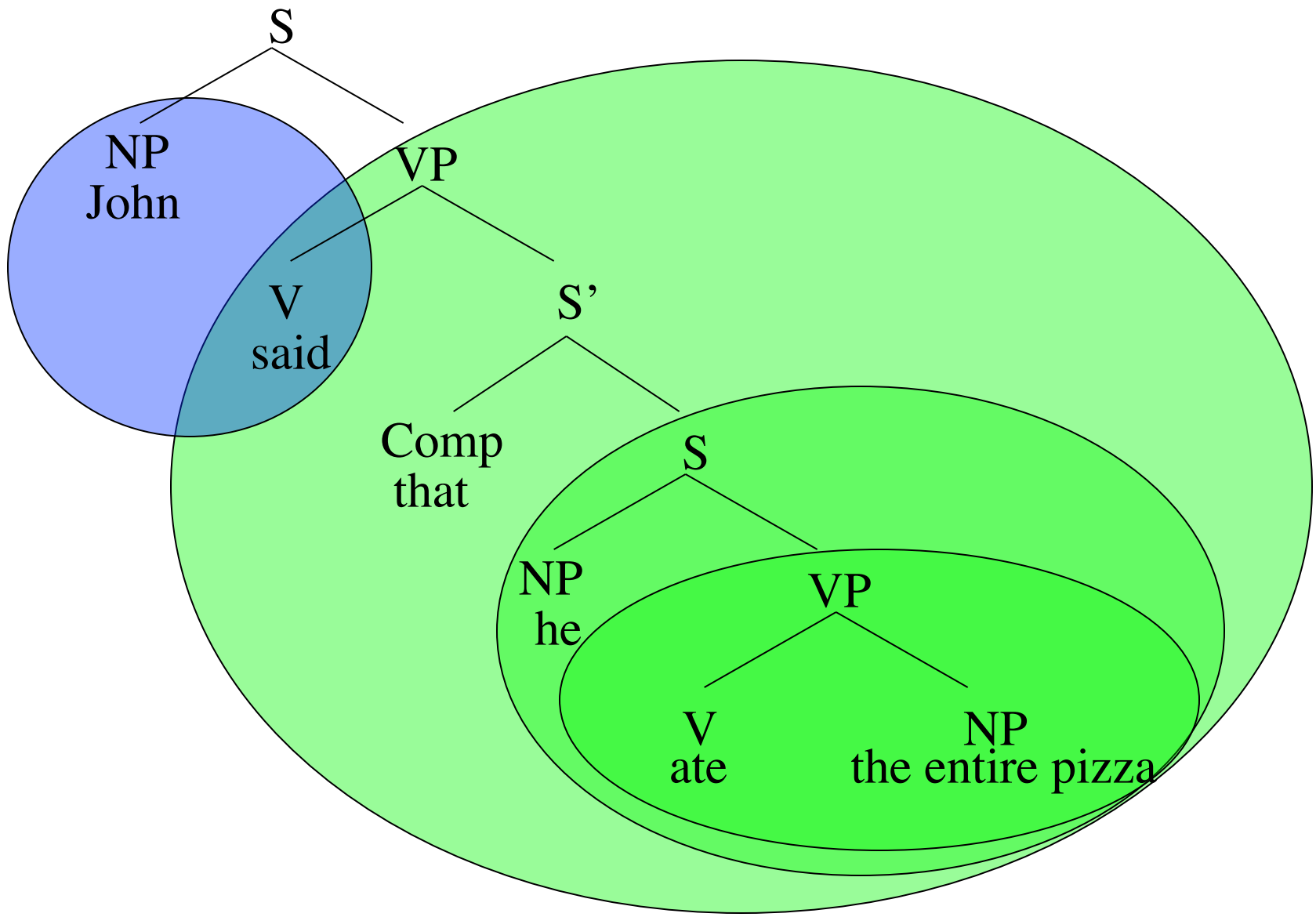
resource limitations

memory architecture and access

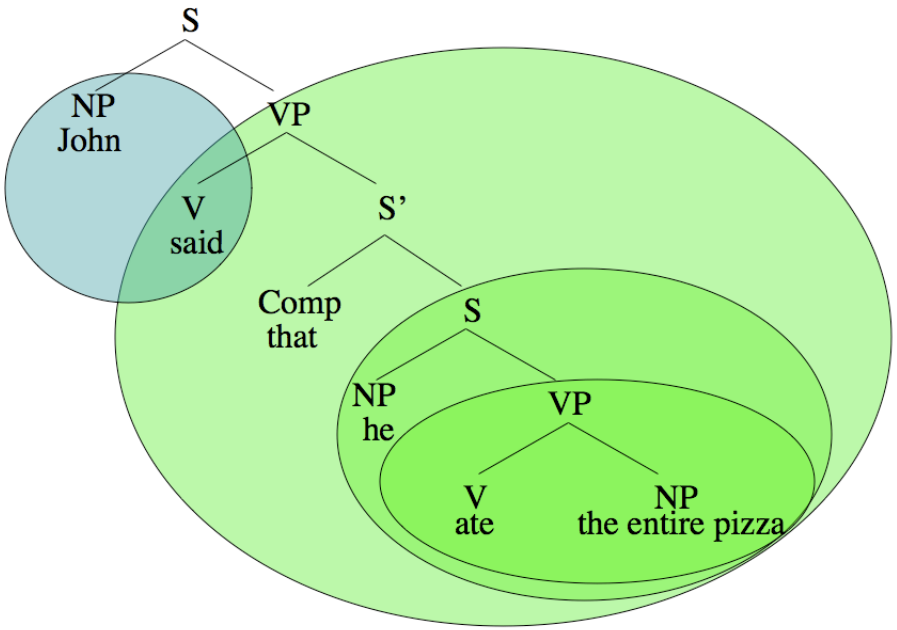


non-discrete units

etc.



An order-based account: Phillips 1996, 2003



- Many tools used to diagnose groupings of words:
  - coordination
  - deletion
  - interpretation (coreference, scope)
  - movement, focus, topicalization
  - etc.

- There are many cases where the tools converge on the same result
- There are also many cases where the tools yield conflicting results

- Suggestion: this well-known dilemma reflects **opacity effects** arising from left-right derivations

  (cf. classic arguments for derivations in phonology)

An order-based account: Phillips 1996, 2003

Immediate Constraint Application

- **Anaphora: Principle C**
  English: Kazanina et al. 2007; Japanese: Aoshima et al. 2009;
  Russian: Kazanina & Phillips 2010
- **Anaphora: Principle A**
  Nicol & Swinney 1989; Clifton et al., 1998; Badecker & Straub 2002; Sturt 2003; Xiang et al. 2009; Dillon et al. 2012
- **Islands**
  Subjects: Stowe 1986; Bourdages 1992; Traxler & Pickering 1996; Omaki & Schulz 2010; Rel. Cl.: McElree & Griffith 1998; CSC: Wagers & Phillips 2009; wh-islands: Neville et al. 1991; Japanese: Yoshida et al. 2004; Parasitic gaps: Phillips 2006
- **Sluicing**
  Yoshida et al. 2011
- **Agreement** (selective)
  Wagers, Lau, & Phillips 2009
- **Case** (selective)
  Bader & Bayer 2006
- **Anaphora: Principle B** (disputed)
  Nicol & Swinney 1989; Clifton et al. 1997; Runner et al. 2006; Lewis, Chow, & Phillips 2012; BUT: Badecker & Straub 2002; Lee & Williams 2006

In ERP studies, rapid error detection is so commonplace that it is news if we fail to find detection of an anomaly within a few tenths of a second!

Reasons for mismatches

- Distinct systems (covering grammar, heuristics/strategies)
- One system, noisy architecture

e.g., NVN = clause
e.g., error-reversal system

- One system, access to internal stages of computation
- One system: mischaracterized grammar
- One system: semantic garden paths

Is the parser working *too well*?

- Simple constraint: pronouns disallow clause-mate antecedents

  Dianne proclaimed that Janet outdid her.

  Binding Principle B: A pronoun must be locally free. (Chomsky 1981)

- Not so simple! Exceptional cases reveal more complex syn-sem interactions

  You know what Mary, Sue, and John have in common? Mary admires John, Sue admires John, and even John admires him, too.

  Everybody hates Larry. Only he himself pities him.

  Exceptions apply to coreference, but not to bound variable use of pronouns

  → motivates claim that 'Janet outdid her' is syntactically fine
  but not 'Everybody outdid her'. Complex interaction of syntax/pragmatics

Is the parser working *too well*?

- Dramatic developmental evidence emerged in late 1980s (Chien & Wexler '90)

Children distinguish	*Mama Bear washed her.*	many errors
	Every Bear washed her.	few errors

- The complex account of Principle B became mainstream
  (… though innocent youth in intro courses largely spared)

- But we inadvertently undermined this elegant finding
  (Conroy, Lidz, Takahashi, & Phillips, 2009)

  … so much for the developmental dissociation

"Grumpy painted him"

Is the parser working *too well*?

- Perhaps adults would show the dissociation (Lewis, Chow, Phillips 2012)

  The actor revealed that the producer had doubted him …
  The actor revealed that every producer had doubted him …

- And the grammatical/pragmatic account predicts fleeting consideration of unacceptable antecedents

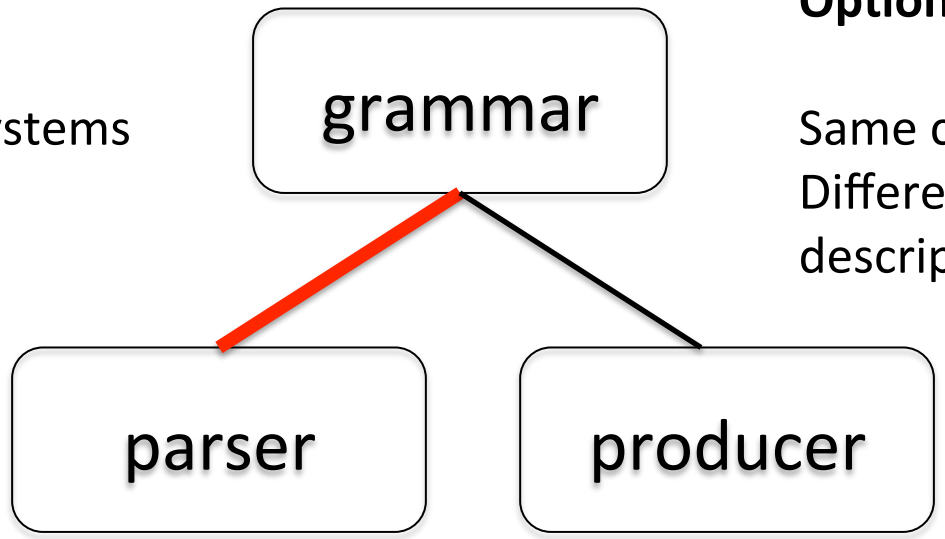
- Fail! No referential/quantificational contrast. No illicit antecedents.
- x7

- In this case the parser's access to antecedents is surprisingly restrictive (but cf. Badecker & Straub 2002)

- Could reflect a simpler covering grammar.
  Or perhaps the Syntax 1 account of Principle B is not so bad after all

Option 1

Distinct cognitive systems



Option 2

Same cognitive system;
Different levels of description (cf. Marr 1982)

To the extent that Fast Stuff and Slow Stuff align closely …

… it minimally motivates a tight link
… it removes a motivation for the division
… but always hard to argue against a clever covering grammar

Motivations for this distinction are different.

As useful as ever to give abstract, process-neutral descriptions
An implementation dependent abstraction is a convenience
Its process-neutrality doesn't mean that there's real process-neutrality

Finally: Capturing the Alignment

- On-line findings motivate mechanisms that apply grammatical constraints highly incrementally (modulo Mark Steedman's important caveat about 'cheating')

- Most grammar formalisms allow this (… or can be adjusted to do so)

- That might not be enough. Evidence suggests that specific order may be forced

  … which is more surprising under accounts that offer flexible combinatorics

- Not today's focus: Forced left-right structure building may provide better answer to Steedman's constituency problem – which phenomena target 'odd constituents', and which ones do not (Phillips 1996, 2003)

- Digging deeply into source of mismatches is important going forward
- Understanding Parser-Producer relation is crucial for Parser-Grammar relation
- Psycholinguists could also help to develop process models of Slow Stuff