Variability in Semantic Cue Effectiveness on Syntactic Ambiguity Resolution: Inducing Low-Span Performance in High-Span Readers

Edward C. Eastwick and Colin Phillips
University of Delaware
tbear@udel.edu colin@udel.edu

0. Introduction

Syntactic ambiguity resolution is affected by non-syntactic constraints, including semantic and pragmatic constraints. The ability to use these constraints varies from individual to individual.

- Goal: Test the effect of increased syntactic complexity on the use of semantic constraints.
- The Capacity Theory of Comprehension (Just & Carpenter, 1992) predicts that availability of semantic cues will decrease with an increase in demands for syntactic processing.
- A syntactic choice model (Grodner, Gibson, & Tunstall, 1999) predicts that semantic cues will have an effect only when syntax has not made a clear choice
- Our claim is that competition for resources can be demonstrated using a within-subjects manipulation.

Main Verb/Reduced Relative (MVRR) ambiguity

(1) The defendant examined by the lawyer was not very reliable.

At the verb examined, there are two possible readings:
"The defendant examined …..
a) Reduced Relative √ …… by the lawyer was not very reliable."
b) Main Verb √ …… the document very closely."

Syntactic disambiguation in (1) occurs immediately, at the word "by".

- Before this, both structures are *plausible*
- We should see a garden-path at the by-phrase

(2) The evidence examined by the lawyer was not very reliable.

"Evidence" allows a semantic disambiguation *before* the "by"-phrase:
"The evidence examined …..
a) Reduced Relative √ …… by the lawyer was not very reliable."
b) Main Verb ??? …… the document very closely."

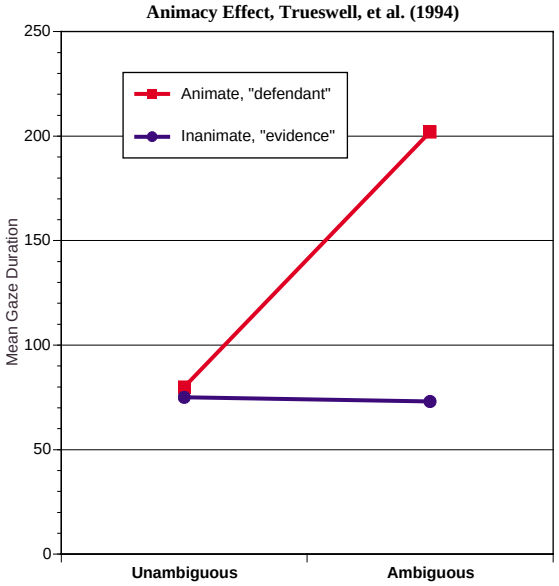
- garden-path effect

Processing at the "by"-phrase should be easier when given the early semantic cue for disambiguation.
Subjects are able to use a variety of non-syntactic cues to aid in the processing of the ambiguity (see Spivey-Knowlton, et al., 1993; Spivey-Knowlton & Sedivy, 1995; Trueswell, et al., 1993; Gibson & Pearlmutter, 1998)

1. Evidence for immediate use of semantic cues

Trueswell, Tanenhaus, & Garnsey (1994)
Eye-tracking experiment using MVRR ambiguity. Subjects read ambiguities headed by animates and inanimates, as well as unambiguous controls.

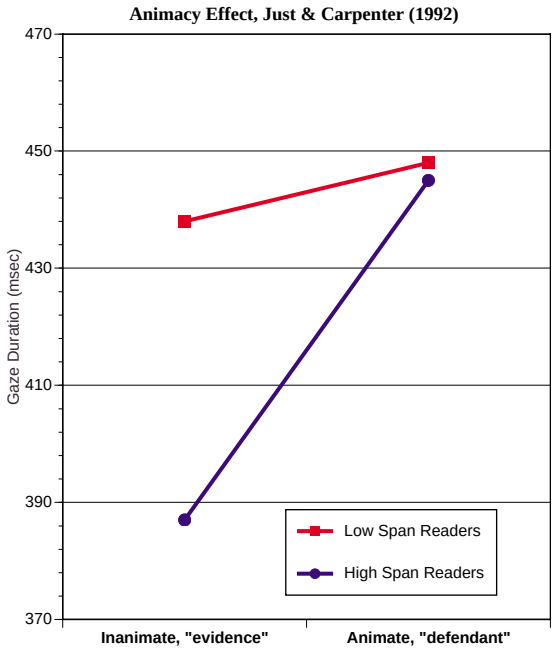
Results
Processing slowdown at the disambiguating "by"-phrase *only* in the ambiguous condition with an animate noun ("the defendant examined"), as compared to unambiguous controls:



2. Causes of variability in semantic cue effectiveness

Just & Carpenter (1992): Difference between individuals
Main Verb/Reduced Relative ambiguities similar to Trueswell, et al. (1994)
Divided subjects into Low and High Reading Span groups (Daneman & Carpenter, 1980). The Reading Span task requires subjects to hold a series of words in memory while they perform a sentence comprehension task.

Results
Only High Span subjects showed evidence that they were able to use the semantic cue of animacy.
Low Span subjects showed no difference in reading times to sentences with Animate vs. Inanimate nouns.



Other research has demonstrated Reading Span differences in sentence processing tasks (e.g. MacDonald, et al., 1992; Pearlmutter & MacDonald, 1995; Friederici, et al., 1998)

Grodner, Gibson, & Tunstall (1999)

Can manipulation of the syntactic environment affect the availability of semantic cues?

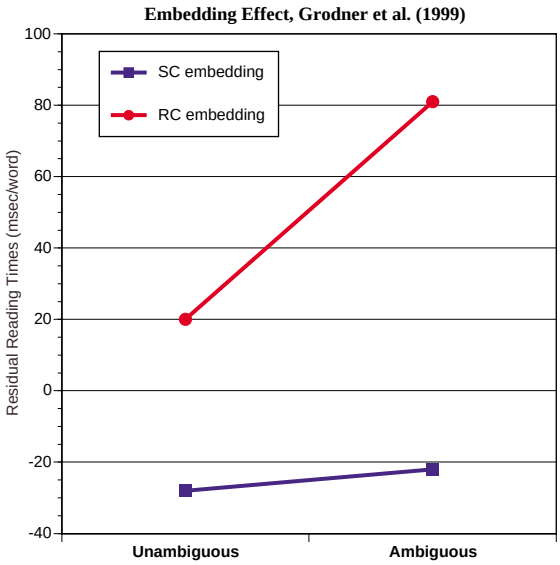
The claim is that semantic cues are used only when syntax does not pick a clear winner. There is no resource limitation on the use of non-syntactic information; it's use is governed only by the syntactic choices available

MVRR ambiguity embedded in different syntactic environments with different syntactic choices:

> *Simple: sentential complement* (SC)
> The witness thought **that** the evidence examined by the lawyer implicated his next-door neighbor.
> *Complex: relative clause* (RC)
> The witness **who** the evidence examined by the lawyer implicated seemed to be very nervous

Results

Evidence for use of animacy cue only in the simpler syntactic condition (SC). No evidence for use in the RC condition.



These results are not inconsistent with theories based on global resource capacity. The RC condition is more complex syntactically (at the point when the ambiguity is being processed the reader is still holding on to the wh-dependency), and therefore resources may not be available to semantic processes.

Other explanations for differences in the use of non-syntactic information

Waters & Caplan (1996)
Measures such as the Reading Span Test do not tap the same types of processes as those which are used for automatic linguistic processing. The processes devoted to syntactic and semantic processing are independent and do not rely on the same pool of resources

MacDonald & Christiansen (1999)
Variability in cue effectiveness is due to variation in linguistic experience – see more extensive discussion below.

3. Current Study

Three self-paced reading experiments – aiming to test a more minimal contrast than found in studies reviewed above. One neural network modeling study.

- within-subjects design
- local syntactic choices for the ambiguity do not change
- varying resource demands

Experiments 1 & 2

See discussion below. [The results were very similar to Experiment 3, but the earlier studies contained fewer controls, so let's move directly to …]

Experiment 3

The claim (Just & Carpenter, 1992): "The larger [resource] capacity of some individuals permits interaction among syntactic and pragmatic information" (p.122).
Low and High Span subjects share the same basic parsing architecture, but competition for resources limits the processing effectiveness for non-syntactic cues in Low Span readers.

The prediction: It should be possible to increase the syntactic load of High Span subjects to the point that they can still process syntax, but are no longer able to use semantic cues, mimicking the performance of Low Span readers. This can be done without changing the local syntactic choices that are available at the point of ambiguity.

Subjects: 56 University of Delaware students (8 additional subjects excluded for scoring <80% on comprehension
questions to filler items and/or <75% on ambiguous test items)
Method: word-by-word, self-paced Moving Window paradigm; all items followed by a Yes/No comprehension
question. Stimuli were presented in two blocks of 100 sentences each.
Stimuli: 2 X 2 X 2 design: ambiguity (ambiguous, unambiguous), animacy (animate noun, inanimate noun),
embedding (embedded, unembedded); 48 test items in 8 conditions, 46 distractors, 106 fillers (stimuli from other experiments). Test stimuli included slightly modified test sentences from Trueswell, et al. (1994)

- Simple (Unembedded) test items:
  Ambiguous, Inanimate
  The evidence recently examined by the lawyer was not very reliable.
  Ambiguous, Animate
  The defendant recently examined by the lawyer was not very reliable.
  Unambiguous, Inanimate
  The evidence that was recently examined by the lawyer was not very reliable.
  Unambiguous, Animate
  The defendant that was recently examined by the lawyer was not very reliable.

- Embedded test items
  Simple test items, as above, embedded in a sentence, interrupting a **subject/verb** dependency:

(4) The judge remembered that the **memo** stating that [*insert simple sentence here*] **had been stolen** from the filing cabinet.

Notes on stimuli:
- An adverb was inserted before the ambiguous verb in each test sentence; these adverbs were mostly adverbs of time, and did not bias the ambiguity in favor of an agentive, main verb reading. This was done due to the fact that in the unambiguous sentences, the overt auxiliary

"was" is a very strong cue that a verb will follow. The adverb equalizes the cloze probability for seeing a verb in the ambiguous and unambiguous conditions. This corrects a confound seen in Experiments 1 & 2

- Embedded sentences within experimentally defined limits of comprehensibility at the ambiguous regions (Gibson, 1999, DLT model of sentence processing)

Reading Span Test: all subjects participated in a version of the Daneman & Carpenter (1980) reading span test based on Roberts, et al. (1999). Very similar to Daneman & Carpenter, except that comprehension questions are asked of all sentences. This makes the task harder and lowers the overall scores

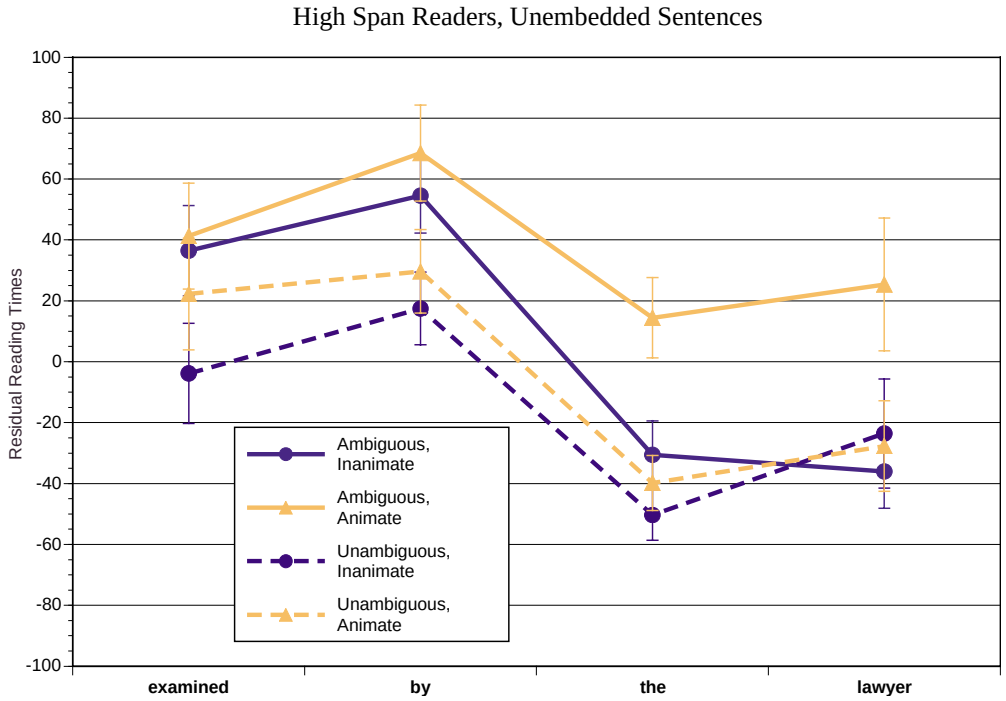
High Span ≥ 2.4
Low Span < 2.4

Analyses: all analyses based on residual reading times; trials in which subjects answered the comprehension questions incorrectly were excluded. Residual reading times trimmed to 4 s.d. (affects 2% of all stimuli)

Results

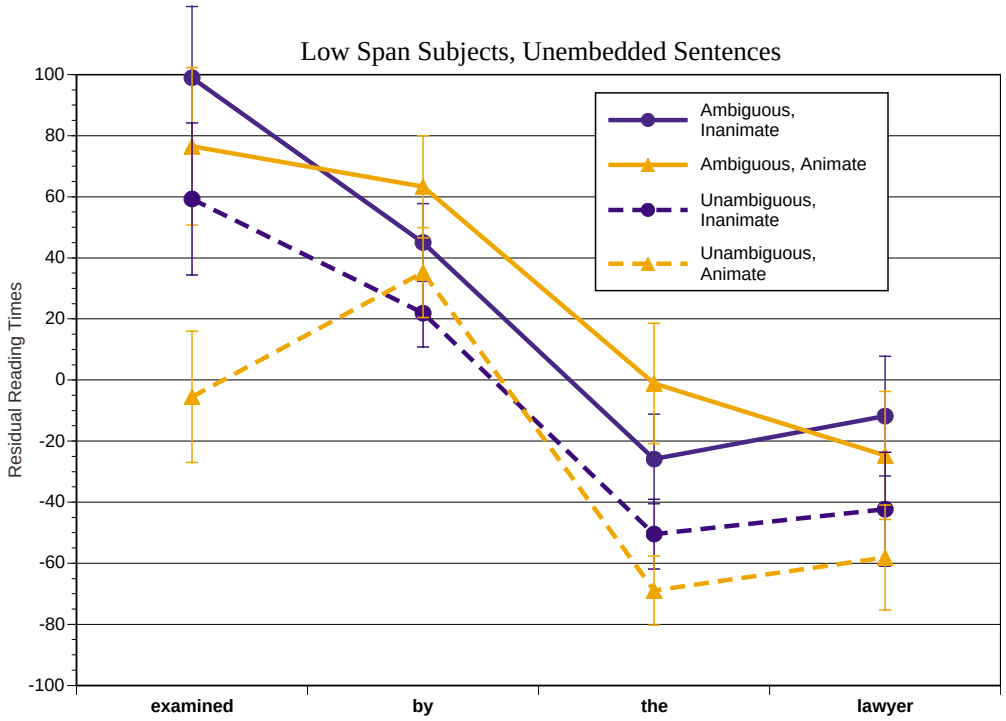
Comprehension question accuracy did not differ significantly by groups:

High Span, Unembedded:	92.8%
Low Span, Unembedded:	89.5%
High Span, Embedded:	82.5%
Low Span, Embedded:	81.7%



High Span, Unembedded (Simple) Conditions:

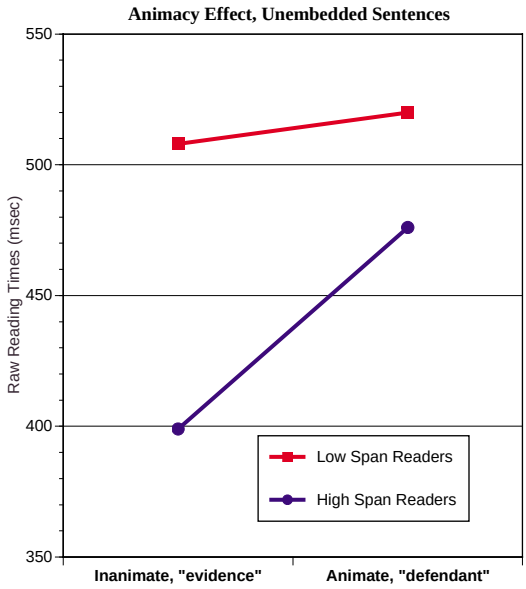
- High Span subjects show evidence that they are using the semantic cue to help guide structure (replicates finding of Trueswell, et al. (1994), and High Span result from Just & Carpenter (1992)).
- Main effect of animacy ($F_1(1,28)=13.11$, $p<0.001$; $F_2(1,47)=5.35$, $p<0.05$) and interaction (animacy x ambiguity, $F_1(1,28)=6.69$, $p<0.01$; $F_2(1,47)=6.94$, $p<0.01$) in the region following "by" ("the lawyer")



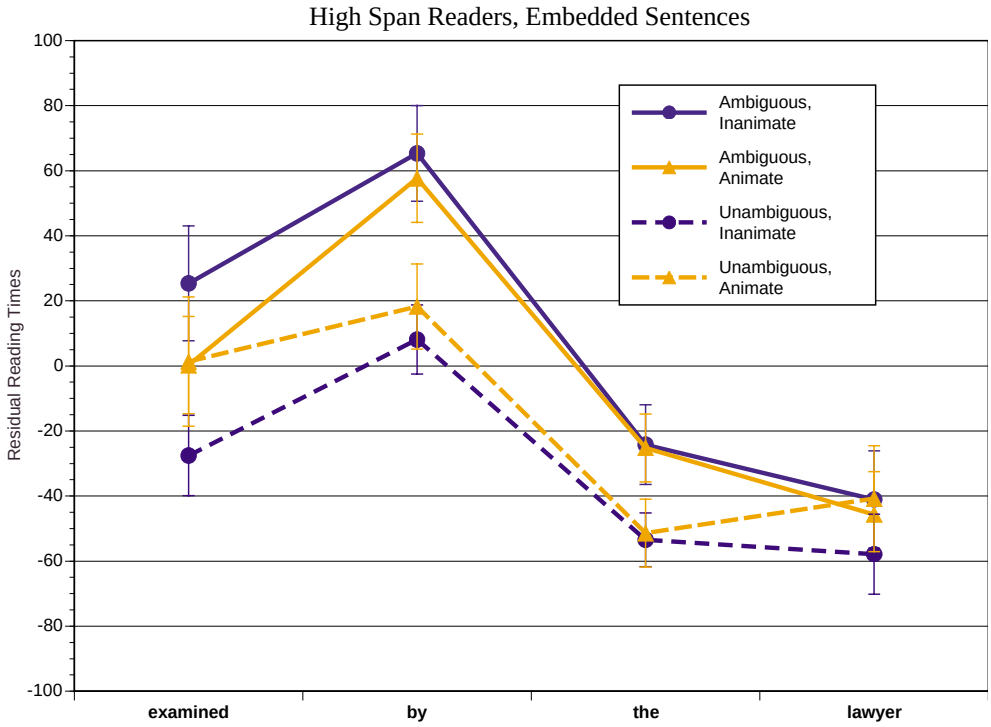
Low Span, Unembedded (Simple) Conditions:

- No effect of animacy or interaction with ambiguity in the region following "by";
- Main effect of ambiguity at "the lawyer" ($F_1(1,26)=11.64$, $p<0.001$; $F_2(1,47)=6.66$, $p<0.01$) and at "examined";
- The lack of any effect of animacy for the Low Span subjects shows that they are not able to integrate the semantic cue to help in parsing.

Unembedded Sentence Findings: Replicate Just & Carpenter's (1992) finding that High Span subjects are
able to use non-syntactic cues to aid in parsing, while their Low Span counterparts are not able to do so.

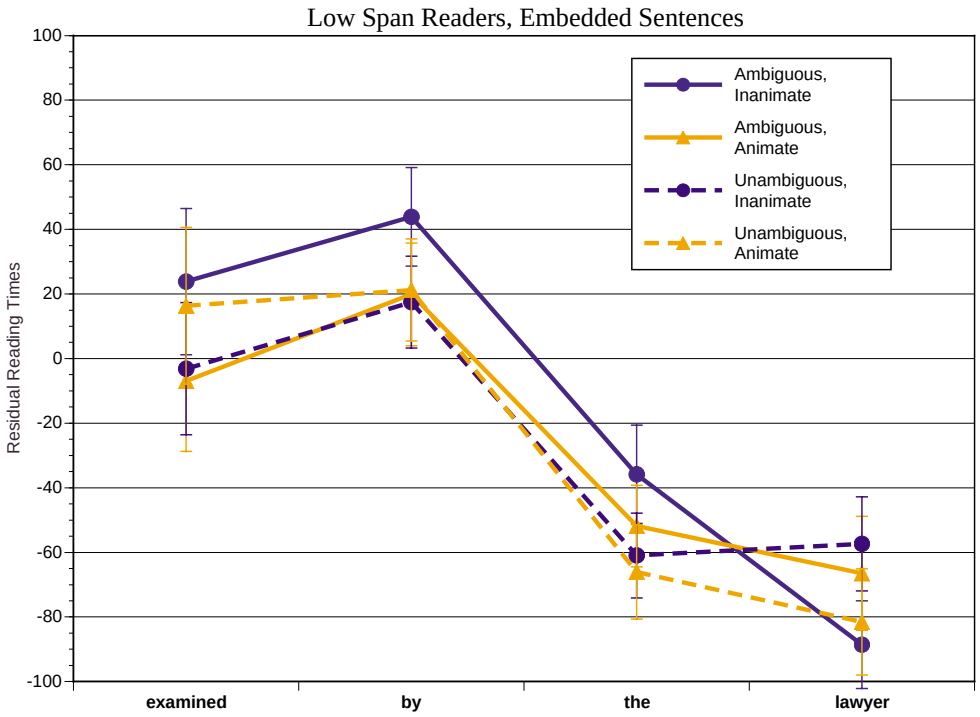


Embedded Conditions



High Span, Embedded Conditions:

- Contrary to the findings from the unembedded conditions, High Span subjects are not sensitive to the animacy cue when the ambiguity occurs in an overall syntactically complex environment where the *local* syntactic choice has not changed
- No effect of animacy or interaction with ambiguity in the regions following "by"; Main effect of ambiguity starting at "by" ($F_1(1,28)=12.78$, $p<0.001$; $F_2(1,47)=9.79$, $p<0.005$)

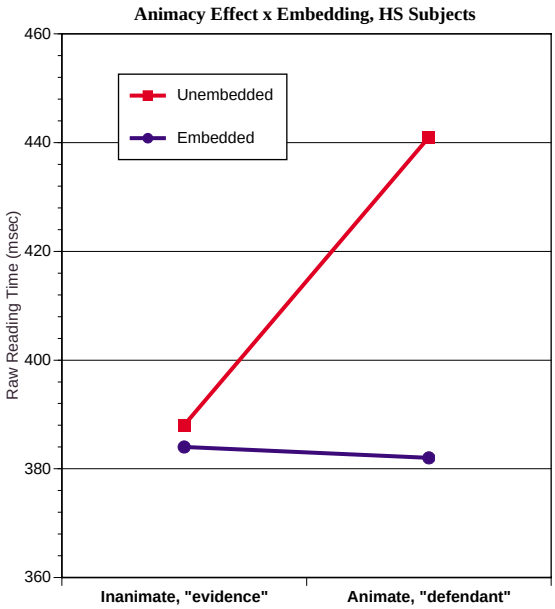


Low Span, Embedded Conditions:

- No effect of animacy or interaction with ambiguity; no main effects of ambiguity.
- Low Span subjects are (not surprisingly) still unable to use the semantic cue when the syntactic environment is very complex.

Embedded Sentence Findings

- High Span, but not Low Span, subjects able to use animacy cues in the unembedded condition.
- When embedded in a complex syntactic environment, High Span subjects lose this ability. We have replicated the Just & Carpenter (1992) result *within-subjects* for the High Span group; they avoid the garden-path by using the animacy cue



- When given a heavy syntactic processing load, High Span subjects are no longer able to utilize non-syntactic cues to constrain their structure-building.

(Note: This graph is slightly different from the Just & Carpenter results, in which the Low Span subjects were slower overall. This effect is due to the fact that Low Span subjects have slower overall reading times; when plotting the effect of embedding for High Span subjects, we see that they read more slowly *only* in the unembedded, ambiguous, animate condition, where a slowdown is expected.)

4. Experiments 1 & 2

- These results are supported by an two earlier versions of this same experiment; in the first (48 UD undergraduates) only High Span subjects were able to use semantic cues in unembedded conditions. The experiment was revised because stimuli in this earlier experiment did not contain an adverb before the ambiguous verb, which led to large main effect of ambiguity already at the main verb, a region earlier than would be expected. This experiment also did not include the unambiguous controls for the unembedded ambiguous sentences.
- This earlier version was also run on undergraduates at MIT , but they were not given the Reading Span task; as a group, the MIT students all behaved like High Span subjects at the University of Delaware (thanks to Ted Gibson & San Tunstall for their assistance in running this experiment)

- All three studies show the same basic finding: the animacy cue is effective in simple sentences for those who can take advantage of it. This advantage disappears when the ambiguities appear in embedded contexts.

5. Conclusion:

Confirmation that:

> (a) non-syntactic cues can have an immediate disambiguating effect in simple ambiguous MV/RR sentences.
> (b) the ability to utilize these cues varies among individuals, according to the amount of leftover resources available to compute non-syntactic information.

New evidence that:

> (a) subjects who are able to use a semantic animacy cue to disambiguate simple Main Verb/Reduced Relative ambiguities lose this ability when the ambiguities are embedded in a complex syntactic environment.
> (b) syntactic and semantic processes compete for the same pool of resources; it is possible to see the effect of this competition in a within-subjects manipulation of syntactic environment, where the *local syntactic choices remained the same*.

6. Network Simulations

Another possible explanation that has been brought to our attention comes from network simulations involving replications of Reading Span differences.

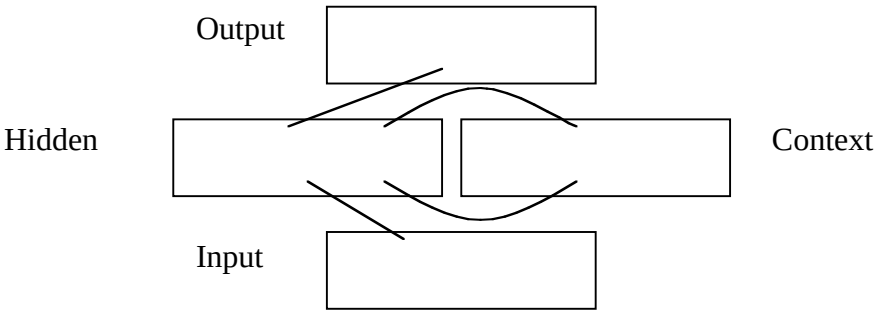
MacDonald & Christiansen (1999)
M & C's claim is that differences in Reading Span are not due to differences in underlying processing architecture. These differences are caused by different amounts of exposure to complex sentence structures. LS subjects have less experience with MVRR ambiguities, and so are unable to deal with them efficiently. HS subjects have had this experience and are well-equipped to deal with these structures. However, even HS readers might have trouble with very complex sentences which are ultra-rare.
M & C. model these scenarios using *simple recurrent network* connectionist models (Elman 1991, 1993).

Possible analysis of our experimental results: high-span subjects follow the main verb parse in embedded contexts due to the extreme rarity of these constructions, *which causes readers to revert to a more 'general' strategy/pattern*.

Modeling question: A (probabilistic) neural network model which has been given very little (or no) evidence in its training set for embedded ambiguities could respond to extremely rare sentences in either of two ways:
(a) Generalization based on global statistical cues:
(b) Generalization based on local statistical cues:
We are testing to see which direction a simple recurrent network will favor. We are not aiming to replicate the results from our experiment. We are trying to build a network which can simulate the Trueswell et al. (1994) finding in simple contexts and then test how it performs on RRs in a novel, embedded context in which it has previously only seen MVs.

Network simulation: designed to test whether or not a network will extend generalizations from simple contexts to embedded contexts containing the same exact local information.

Network architecture: 31 input nodes
31 output nodes
30 hidden units
30 context units



Training stimuli:

- 24 unique input nodes: 4 animate nouns, 5 inanimate nouns (one not used), 12 verbs, 'that', 'by', 'stop'
- 7 category nodes: noun, verb, 'that', 'by', 'stop', animate noun, inanimate noun
- 3500 sentences from four conditions
  1500 simple MV: [*animate* noun] - [verb] - [animate/inanimate noun]
  "The defendant examined the evidence/student."
  1000 embedded MV: [animate noun] – [verb] – ['that'] - [*animate* noun] - [verb] - [animate/inanimate noun]
  "The witness thought that the defendant examined the evidence/student
  500 simple animate RR: [*animate* noun] – [verb] – ['by'] – [animate noun] – [verb]
  "The defendant examined by the witness lied."
  500 simple inanimate RR: [*inanimate* noun] – [verb] – ['by'] – [animate noun] – [verb]
  "The evidence examined by the witness was unreliable."
- Network trained on 5 epochs (17,500 sentences); the error calculation for the network reached an asymptotic level after half of those sweeps.

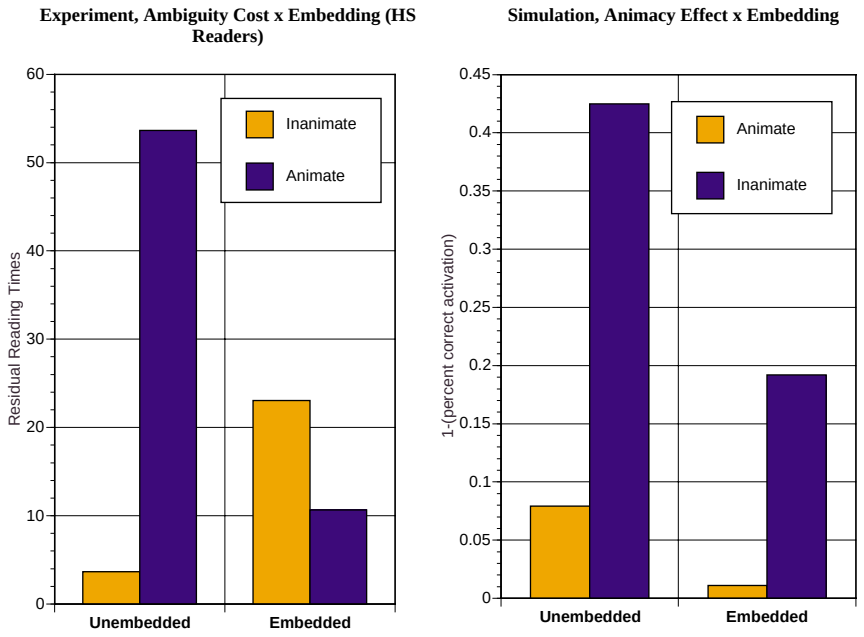
Testing:

- 60 sentences, 10 from each condition in the training set plus 10 from 2 novel conditions:
  embedded animate RR: [animate noun] – [verb] – ['that'] - [*animate* noun] – [verb] – ['by'] [animatenoun] – [verb]
  "The judge knew that the defendant examined by the witness lied."
  embedded inanimate RR: [animate noun] – [verb] – ['that'] - [*inanimate* noun] – [verb] – ['by'] - [animate noun] – [verb]
  "The judge knew that the evidence examined by the witness was unreliable."

Scoring: The network was tested on a set of sentences containing main verbs and reduced relatives, in both simple and embedded contexts. Note that the network was not given any RRs

in the input. Network performance on these test sentences is measured by a proportion of activation dedicated to predicting the correct category of the next word in the sentence. Since what the network really had to learn was the categories (for example, when to expect a 'by' given an inanimate noun – verb sequence), analyses were carried out on the category nodes only. This actually should increase the accuracy of the system, since it does not have to predict which individual *token* is coming next, only the category of that token. The category predictions are much more consistent throughout the data.

Experiment Data vs. Simulation Prediction Accuracy



Important Note on above graph: The measures used for reading times and prediction are different. For reading times, *poor* performance is seen as an increase in reading time. The animacy effect we are looking at is a relative difficulty with animate nouns compared to inanimate nouns. The simulation measure is one of *good* performance; the same animacy difference would produce an increase in prediction scores for *inanimates*, relative to animates

- The network replicates the animacy effect in the simple conditions, replicating the Trueswell et al. (1994) finding. However, in the embedded conditions, it still shows an effect of animacy, which is *not* the result that we saw in the embedded conditions of the experiments on humans.

- While our simulation may not be the best model to compare with the results from humans, it at least shows that the network will make predictions based on *local* information when it is presented with environments that it has never seen before. This means that the simple recurrent network does not provide an immediate explanation for the failure of semantic cues in syntactically complex environments.

References:

Daneman, M. and P.A. Carpenter (1980). "Individual differences in working memory and reading", in *Journal of Verbal Learning and Verbal Behaviour*19, 450-466.

Christiansen, M. H., and N. Chater (in press). "Toward a connectionist model of recursion in human linguistic performance", *Cognitive Science*

Elman, J.L. (1990). "Finding structure in time", *Cognitive Science*, **14**, 179-211

Elman, J.L. (1993). *Rethinking Innateness*, Cambridge, MA, MIT Press

Friederici, Angela D., K. Steinhauer, A. Mecklinger, and M. Meyer (1998). "Working memory constraints on syntactic ambiguity resolution as revealed by electrical brain responses", *Biological Psychology*, 47(3), 193-221.

Gibson, Edward (1998). "Linguistic complexity: locality of syntactic dependencies", *Cognition*, 68(1), 1-76.

Gibson, Edward, and Neal Pearlmutter (1998). "Constraints on sentence comprehension", *Trends in Cognitive Sciences*, 2(7), 262-268.

Grodner, Daniel, Edward Gibson, and SusanneTunstall (1999). "Syntactic complexity in ambiguity resolution." ms, MIT

Just, Marcel Adam and Patricia A. Carpenter (1992). "A Capacity Theory of comprehension: Individual differences in working memory", *Psychological Review* 99(1), 122-149.

MacDonald, Maryellen C., Marcel Adam Just, andPatricia A. Carpenter (1992). "Working memory constraints on the processing of syntactic ambiguity", *Cognitive Psychology* 24, 56-98.

MacDonald, Maryellen C., and M. Christiansen (1999). "Reassessing working memory: A reply to Just & Carpenter and Waters & Caplan", *Psychological Review*, submitted

Pearlmutter, Neal J. and Maryellen C. MacDonald (1995). "Individual differences and probabilistic constraints in syntactic ambiguity resolution", *Journal of Memory and Language* 34, 521-542.

Roberts, Rose, Christopher Hooker, and Edward Gibson (1997). "Sentence memory capacity and working memory", Poster presented at CUNY Conference on Human Sentence Processing, Santa Monica, CA March 20-22.

Roberts, Rose, and Edward Gibson (1999). "Individual differences in sentence memory", ms, MIT

Spivey-Knowlton, Michael J., & J Sedivy (1995). "Resolving ambiguities with multiple constraints", *Cognition*, 55(3), 227-267.

Spivey-Knowlton, Michael J., John C. Trueswell, and Michael K. Tanenhaus (1993). "Context effects in syntactic ambiguity resolution: Discourse and semantic influences in parsing reduced relative clauses", *Canadian Journal of Experimental Psychology* 47(2), 276-309.

Trueswell, John C., Michael K. Tanenhaus, and Susan M. Garnsey (1994). "Semantic influences on parsing: Use of thematic role information in syntactic ambiguity resolution", *Journal of Memory and Language* 33, 285-318.

Trueswell, John C., Michael K. Tanenhaus, and Christopher Kello (1993). "Verb-specific constraints in sentence processing: Separating effects of lexical preference from garden-paths", *Journal of Experimental Psychology: Learning, Memory, and Cognition* 19(3), 528-553.

Edward C. Eastwick, Colin Phillips
Department of Linguistics
University of Delaware
46 E. Delaware Ave.
Newark, DE 19716

tbear@udel.edu
colin@udel.edu www.ling.udel.edu/colin